

Semi-Parametric Drift and Diffusion Estimation for Multiscale Diffusions

S. Krumscheid[†] G. A. Pavliotis^{†§¶} S. Kalliadasis[‡]

June 11, 2012

Abstract

We consider the problem of statistical inference for the effective dynamics of multiscale diffusion processes with (at least) two widely separated characteristic time scales. More precisely, we seek to determine parameters in the effective equation describing the dynamics on the longer diffusive time scale, i.e. in a homogenization framework. We examine the case where both the drift and the diffusion coefficients in the effective dynamics are space-dependent and depend on multiple unknown parameters. It is known that classical estimators, such as Maximum Likelihood and Quadratic Variation of the Path Estimators, fail to obtain reasonable estimates for parameters in the effective dynamics when based on observations of the underlying multiscale diffusion. We propose a novel algorithm for estimating both the drift and diffusion coefficients in the effective dynamics based on a semi-parametric framework. We demonstrate by means of extensive numerical simulations of a number of selected examples that the algorithm performs well when applied to data from a multiscale diffusion. These examples also illustrate that the algorithm can be used effectively to obtain accurate and unbiased estimates.

Keywords. Parameter estimation, multiscale diffusions, stochastic differential equations, homogenization, coarse-graining, effective dynamics

AMS subject classifications. 60H10, 60J60, 62M05, 34E13, 74Q99, 62F99

1 Introduction

Problems with multiple temporal and/or spatial scales emerge naturally in a wide variety of fields in science and engineering. From biological phenomena [8] and atmosphere and ocean science [40] to molecular dynamics [21], material science [15], fluid and solid mechanics [24, 26]. Many of such systems are often subject to noise that is either due to thermal fluctuations [13], randomness in the environment (e.g. uncertainty in some parameters) [25, 52, 50], coarse-graining of high-dimensional deterministic systems with random initial conditions [41, 56], or stochastic parameterization of small scale effects [12].

Mathematically the influence of noise in a system can be described by a single or coupled, nonlinear stochastic differential equations (SDEs) with multiple scales. In many cases these equations are

[†]Department of Mathematics, Imperial College London, South Kensington, London SW7 2AZ, United Kingdom: {s.krumscheid10, g.pavliotis}@imperial.ac.uk

[‡]Department of Department of Chemical Engineering, Imperial College London, South Kensington, London SW7 2AZ, United Kingdom: s.kalliadasis@imperial.ac.uk

[§]INRIA MicMac Project Team, Ecole des Ponts ParisTech, 6 et 8 Avenue Blaise Pascal, Cité Descartes - Champs sur Marne, 77455 Marne la Vallée Cedex 2, France: grigorios.pavliotis@cermics.enpc.fr

[¶]Institut für Mathematik, Freie Universität Berlin, Arnimallee 6, 14195 Berlin, Germany: gpavlo@zedat.fu-berlin.de

high-dimensional but from a practical point of view only the evolution of some components of the solution is of main interest, since they act on a slower scale. It is therefore desirable to approximate the full system by an adequate simplified low-dimensional effective model (coarse-graining) that retains the essential dynamic characteristics of the full system. The effective equation is often amenable to analytical and numerical work. However, usually only the complete multiscale (fast/slow) system is directly observable but not the effective dynamics. Consequently, much research has been undertaken to find accurate approximations to the effective dynamics [19].

In many applications a wealth of data (i.e. observations of the fast/slow system) is often available and it is therefore worthwhile to use these data to determine the effective dynamics. However, extracting the effective dynamics directly from the available data is often not straightforward and it is thus important to adapt techniques from analysis, statistics, and numerical analysis to obtain appropriate coarse-grained models.

This data-driven coarse graining methodology has been applied to relatively simple (stochastic) systems for which a low dimensional coarse-grained equation exists, using techniques such as homogenization and averaging. For example, in [46] Brownian motion in a two-scale potential was studied. Therein, through both rigorous mathematical analysis and numerical simulations, it was shown that the estimation of drift and diffusion coefficients in the coarse-grained model is asymptotically biased when using classical estimators. Furthermore, it was shown that subsampling the available data at an appropriate rate between the two characteristic time scales of the full system is necessary for an accurate estimation of both drift and diffusion coefficients in the coarse-grained model. More general fast/slow systems of SDEs for which a coarse-grained equation exists were studied in [45] where it was shown that the same issue of asymptotically biased estimators persists in the homogenization framework – that is, when considering an effective dynamics on the (longer) diffusive time scale – and that appropriately subsampled data reduce the bias. These techniques were then applied to the problem of estimating eddy diffusivities from noisy Lagrangian observations in [9] where an improved algorithm that combines subsampling with appropriate averaging and variance reduction techniques was proposed and tested. Furthermore, inverse problems for multiscale partial differential equations – a problem closely related to that of parameter estimation – were studied in [44].

Although appropriate subsampling can potentially improve the accuracy of estimators in the context of multiscale diffusions, as it yields unbiased estimators, the question of an optimal subsampling rate (i.e. the rate for which the bias is removed) remains open; e.g. the studies in [46, 45] provide only existence results. Furthermore, the numerical experiments presented in the aforementioned works indicate that the optimal subsampling rate is not only problem dependent, but is also different for different parameters in the same model. Consequently, an optimal subsampling rate is in general unknown. The problem of identifying an optimal subsampling rate for Gaussian processes has been studied in [3, 4]; see also [55] for related work in the context of econometrics.

Related problems have been studied in the context of numerical analysis for SDEs with multiple time scales. In particular, the heterogeneous multiscale method (HMM) [54, 12] is based on the idea of evolving the solution of the low-dimensional coarse-grained equation, when the coefficients in the coarse-grained equation are being evaluated “on the fly” by running short runs of the underlying fast dynamics. Similar ideas have been proposed in the framework of the “equation-free methodology” introduced by Kevrekidis and collaborators; see e.g. [53, 31, 30, 32]. Recently, the HMM methodology has been extended to approximate stochastic partial differential equations with multiple timescales [1]. As such, these techniques can be considered as a hybrid between numerical analysis and statistical inference since the coefficients in the coarse grained SDE are estimated from data that are obtained from short runs of the full dynamics. We emphasize, however, that in the HMM the fast dynamics is assumed to be known, whereas in the aforementioned works on parameter estimation for multiscale diffusions, as well as in the present work no such assumption is

made.

We also mention that the effect of the multiscale structure on the evolution of the coarse-grained probability density using the Fokker-Planck equation (Kolmogorov's forward equation) was studied in [17]. In this study it was shown that when decreasing the spatial discretization in a finite difference approximation the error increases rapidly and that in order to avoid this, it is necessary to improve the accuracy of the estimators of the drift and diffusion coefficients.

In many cases of interest the noise in the coarse-grained equation appears in a multiplicative way. One is thus confronted with the problem of estimating parameters in both the drift and the diffusion coefficients of an SDE of the form

$$dx_t = f(x_t; \vartheta) dt + g(x_t; \theta) dW_t ,$$

with unknown parameters $(\vartheta, \theta)^T \in \Theta$, the set of all feasible parameters. The problem is further complicated by the fact that the parameters ϑ and θ may not be independent as for example in the problem of Brownian motion in a two-scale potential, see Section 3.2.2. In the absence of multiscale effects in the data, or if one assumes that the optimal subsampling rate is known, a combination of the maximum likelihood estimator (MLE) and the quadratic variation of the path (QVP) is the most commonly used estimator in practice; see [51, 36, 37] for background material on classical estimators. The QVP (or a variant thereof) is used to estimate parameters in the diffusion coefficient and, based on these estimates, the MLE is used to obtain estimates for the parameters in the drift. As the MLE is based on an SDE with unit diffusion, one usually transforms the original SDE into an SDE with unit diffusion coefficient by applying Itô's formula to

$$h_\theta(x) = \int_c^x g(u; \theta)^{-1} du , \quad (1)$$

for an arbitrary c in the state space of the process x and replacing θ with the estimator obtained from the QVP for concreteness. For example, when considering

$$g(x; \theta) = \sqrt{\theta_1 + \theta_2 x^2} ,$$

which is the diffusion coefficient in the coarse-grained equation of the stochastic Burgers' equation (see Section 3.2.4), then transformation (1) reads

$$h_\theta(x) = \frac{\ln(\sqrt{\theta_2}x + \sqrt{\theta_1 + \theta_2 x^2}) - \ln(\sqrt{\theta_2}c + \sqrt{\theta_1 + \theta_2 c^2})}{\sqrt{\theta_2}} ,$$

where $\theta = (\theta_1, \theta_2)^T$ has to be replaced by a previously obtained estimator.

Notice that this transformation can be singular when implementing it in practice, so that special care has to be taken when performing numerical simulations. In particular within the MLE framework where a (nonlinear) objective function needs to be maximized, this might cause problems. An alternative approach to estimate parameters is based on the MLE for the discretized approximation, e.g. obtained by the Euler-Maruyama scheme [33, ch. 9.1], of the time-continuous SDE. However, the non-vanishing (fixed) time-step size introduces an additional bias so that even for simple models this approach does not necessarily yield consistent estimators [38].

The main aim of the present study is to develop statistical inference techniques that enable us to estimate parameters in both drift and diffusion coefficients of a coarse-grained equation in the presence of an underlying (either stochastic or deterministic) multiscale structure in the fast/slow system. Furthermore, we wish to extend this approach in a semi-parametric framework, to situations where the drift and diffusion coefficients can be expanded in an appropriate (e.g. Taylor series) expansion.

Besides the lack of reliable statistical inference techniques for these problems, the motivation for this study originates also from recent results on the derivation of coarse-grained equations (also known as amplitude equations) for stochastic partial differential equations (SPDEs) with quadratic nonlinearities [6]. A typical example for such a SPDE is the stochastic Burgers' equation

$$du_t = \left((\partial_x^2 + 1)u_t + \partial_x u_t^2 + \varepsilon^2 u_t \right) dt + \varepsilon \mathcal{Q} d\mathcal{W}_t, \quad (2)$$

on $[0, \pi]$ equipped with appropriate boundary conditions. Therein \mathcal{Q} denotes the covariance operator, \mathcal{W} space-time white noise, and $0 < \varepsilon \ll 1$. To study solutions to (2) of $O(\varepsilon)$ on time scales of $O(\varepsilon^{-2})$, i.e. ensuring that we are in the regime described by amplitude equations, a diffusive rescaling is performed by defining v via $\varepsilon v(\varepsilon^2 t) = u(t)$. Then v solves

$$dv_t = \left(\frac{1}{\varepsilon^2} (\partial_x^2 + 1)v_t + \frac{1}{2\varepsilon} \partial_x v_t^2 + v_t \right) dt + \frac{1}{\varepsilon} \mathcal{Q} d\mathcal{W}_t. \quad (3)$$

If the SPDE is equipped with homogeneous Dirichlet boundary conditions it can be shown that the dominant mode of the solution to (3) can be approximated by the solution to an one-dimensional SDE driven by classical Brownian motion W_t of the form

$$dX_t = (AX_t - BX_t^3) dt + \sqrt{\sigma_a^2 + \sigma_b^2 X_t^2} dW_t.$$

Since this class of SPDEs arises in many different applications – from population biology [25] to fluid dynamics [50, 49] where many data are available – it is of major interest to obtain effective dynamics for the dominant modes of solutions to the SPDE by means of a data-driven coarse graining methodology.

The statistical inference technique we propose here consists of two steps. First, we use the martingale property of the stochastic integral to obtain an equation involving only the drift but not the diffusion coefficient of the SDE. Since the drift might depend on multiple (unknown) parameters, it is generally impossible to obtain the parameters uniquely from a single equation. The main element we employ to overcome this under-determined situation is the often disregarded initial condition. In fact, by varying the initial condition of the SDE one can define the estimator for the drift parameters via the best approximation of a system of equations. For the second step, we rely on the estimators for the drift parameters and on the Itô isometry to obtain a relation for the unknown parameters in the diffusion. Using the same idea as in the first step, that is by varying the initial condition of the SDE, we can also define the estimators for parameters concerning the diffusion via the best approximation of a system of equations. In both steps the estimators rely on many short trajectories (i.e. on ensemble averages) in contrast to classical estimators that rely on a long trajectory (i.e. on ergodicity) of the underlying process, see e.g. [51].

The main practical advantages of the methodology proposed here can be summarized as follows:

- (a) Coarse-grained models of high-dimensional, possibly infinite-dimensional, problems that retain the essential dynamic characteristic of the full system are a powerful tool to study problems arising in science and engineering. The simplicity of such an effective model, typically an SDE, makes it attractive for mathematical and numerical scrutiny. When the coefficients of the SDE can be computed exactly, the effective SDE allows us to decipher rapidly some of the basic characteristics of the full system, e.g. by computing first-passage properties and properties alike as it was done for instance in [50, 49]. However, in several problems it is not straightforward to obtain the coefficients of the coarse-grained SDE exactly due to the complexity of the underlying full system. From a theoretical standpoint several assumptions need to be made, e.g. the dynamics is dominated by a single eigenfunction/mode (often associated with a symmetry in the system) and the higher-order modes decay sufficiently fast

(see the derivation of the “phase-diffusion” equation describing the transverse instability of propagating waves/fronts; e.g. in [35, 28]), to obtain computable coefficients. Conversely, from a practical point of view obtaining approximations (estimators) of the coefficients in the coarse-grained model, which can be used to study the basic characteristics of the full system, is very appealing.

- (b) A model for a physical or technological process might not be readily available, either because its derivation is cumbersome, or it is not straightforward to formulate it from first principles. But the underlying physics of the phenomena at hand and previous experience with similar systems suggests an SDE of the form adopted here, at least in certain regions of the parameter space. In a spirit similar to the equation-free approach, one can utilize available data and obtain a low-dimensional approximation to the process, which in turn can be used as a model for the process in regions of the parameter space consistent with the assumptions imposed from the outset. But often such models can be applicable in regions beyond those dictated by the assumptions. For instance, for a long-wave instability such as that observed on a surface of a film flowing down an inclined plane [29] the growth rate curve of infinitesimal disturbances extends from the origin up to a cut-off wavenumber. The Landau-Stuart equation is then only applicable sufficiently close to the cut-off, i.e. it can be used to describe the transition when an unstable wave motion of given (“fundamental”) wavenumber interacts with its first stable harmonics. On the other hand, when the growth rate curve has a “nose” near criticality consistent with a short-wave instability and so that the system equilibrates to a stationary norm solution in the nonlinear regime, such as with Rayleigh-Bénard convection [10], the Landau-Stuart equation can be applicable even past the instability threshold.

The rest of the paper is organized as follows. In Section 2 we present the precise derivation of the estimators for systems with and without multiscale structure. Based on these results, in Section 3 the general applicability and performance of the proposed methodology is investigated via different numerical examples. The extensive numerical study we undertake illustrates that the proposed technique enables us to estimate accurately parameters in multiscale diffusions. A summary of the results and future perspectives are offered in Section 4.

2 Estimators

We present the precise derivation of the drift and diffusion estimators for systems without and systems with multiscale effects present. First we outline the methodology for SDEs without multiscale structure and illustrate some properties of the estimators in this case, before presenting the set-up for multiscale diffusions.

2.1 Derivation of Drift and Diffusion Estimators

For the sake of simplicity we consider here only one dimensional real-valued processes (see Section 3.2.3 for an example of a multivariate process). Consider the scalar-valued Itô stochastic differential equation

$$dx_t = f(x_t) dt + \sqrt{g(x_t)} dW_t, \quad x(0) \equiv x_0 = \xi, \quad (4)$$

where W denotes standard one-dimensional Brownian motion. Both the drift coefficient f and the diffusion coefficient g are assumed to be sufficiently smooth such that the SDE provides a unique solution for any initial condition $\xi \in \mathbb{R}$ and on any finite time interval; see e.g. [34, sec. 1]. Moreover, we assume that both drift f and diffusion g depend on unknown parameters and the task is to estimate the parameters in f and g from available data. In fact, here we focus on the case

when $f(x)$ and $g(x)$ are polynomials in x of degree $\max\{J_f\}$ and $\max\{J_g\}$, respectively, where $J_f, J_g \subset \mathbb{N}_0$ denote index sets of finite cardinality $p = |J_f|$ and $q = |J_g|$ respectively. The unknown coefficients of the polynomials are $\vartheta \equiv (\vartheta_j)_{j \in J_f} \in \mathbb{R}^p$ and $\theta \equiv (\theta_j)_{j \in J_g} \in \mathbb{R}^q$ respectively, i.e. we consider

$$f(x) \equiv f(x; \vartheta) := \sum_{j \in J_f} \vartheta_j x^j \quad \text{and} \quad g(x) \equiv g(x; \theta) := \sum_{j \in J_g} \theta_j x^j. \quad (5)$$

Consequently, f and g are linear functions in ϑ and θ , respectively. These particular assumptions on the drift f and the diffusion g simplify the notation in what follows and they will lead to a linear system of equations for the estimators of the parameters.

The starting point in the derivation of the estimators is based on the following identities

$$\mathbb{E}(x_t - \xi) = \int_0^t \mathbb{E}(f(x_s)) ds, \quad (6a)$$

$$\mathbb{E}\left((x_t - \xi - \int_0^t f(x_s) ds)^2\right) = \int_0^t \mathbb{E}(g(x_s)) ds, \quad (6b)$$

owing to the martingale property of the stochastic integral and the Itô isometry, respectively, holding for any fixed initial condition ξ . The next step is to incorporate the parameterization of the functions f and g into (6), to identify the functional structure of the relation for the parameters ϑ and θ , respectively. We begin with substituting the parameterization of f into equation (6a), which will yield an estimator for the parameter ϑ , that is present in the drift term alone. Based on this estimator it is possible to proceed similarly with equation (6b) and eventually obtain an estimator for the parameter θ present in the diffusion part.

Substituting ansatz (5) for f into (6a) yields

$$\mathbb{E}(x_t - \xi) = \sum_{j \in J_f} \vartheta_j \int_0^t \mathbb{E}(x_s^j) ds, \quad (7)$$

for a given initial condition ξ . Therein \mathbb{E} denotes the expectation with respect to the Wiener measure and with respect to processes starting at a fixed initial condition ξ . To emphasize this dependency on the initial condition we will use the notation $\mathbb{E} \equiv \mathbb{E}_\xi$. Fix a time $t > 0$ (the question how to choose the final time t will be addressed in Section 3) and define

$$b_1: \mathbb{R} \ni \xi \mapsto b_1(\xi) := \mathbb{E}_\xi(x_t - \xi) \in \mathbb{R}$$

$$a_1: \mathbb{R} \ni \xi \mapsto a_1(\xi) := \left(\int_0^t \mathbb{E}_\xi(x_s^j) ds \right)_{j \in J_f} \in \mathbb{R}^p.$$

With these definitions equation (6a) can be rewritten as

$$a_1(\xi)^T \vartheta = b_1(\xi). \quad (8)$$

The above equation is under-determined for $p > 1$. To derive a well-defined estimator for ϑ , we consider a finite sequence of initial conditions $(\xi_i)_{1 \leq i \leq m}$ with $m \geq p$. Since (8) is valid for each initial condition, this approach yields a system of linear equations

$$A_1 \vartheta = b_1, \quad (9)$$

with $A_1 := (a_1(\xi_i)^T)_{1 \leq i \leq m} \in \mathbb{R}^{m \times p}$ and $b_1 := (b_1(\xi_i))_{1 \leq i \leq m} \in \mathbb{R}^m$. The linear system does not have a unique solution in general (if a solution exists at all). To overcome this shortcoming we

define the solution of the system of linear equations in (9), i.e. the estimator of the drift parameter, to be the *best approximation*

$$\hat{\vartheta} := \arg \min_{s \in \mathcal{S}_1} \|s\|_2^2, \quad \mathcal{S}_1 := \{z \in \mathbb{R}^p: \|A_1 z - b_1\|_2^2 \rightarrow \min\},$$

respectively,

$$\hat{\vartheta} := A_1^+ b_1, \quad (10)$$

with A_1^+ being the Moore-Penrose pseudo-inverse [5]. We note that the estimation of parameters in the drift does not require knowledge of the diffusion coefficient.

Assume now that we have already estimated the parameters in the drift f correctly. Then substituting the ansatz (5) of g into (6b) yields

$$\mathbb{E}\left((x_t - \xi - \int_0^t f(x_s; \hat{\vartheta}) ds)^2\right) = \sum_{j \in J_g} \theta_j \int_0^t \mathbb{E}(x_s^j) ds.$$

Recall again that the expectation is with respect to the Wiener measure and processes starting at ξ , hence $\mathbb{E} \equiv \mathbb{E}_\xi$. To cope with multiple parameters, we follow the same approach as for the drift parameters above. In fact, define here

$$b_2: \mathbb{R} \ni \xi \mapsto b_2(\xi) := \mathbb{E}_\xi\left((x_t - \xi - \int_0^t f(x_s; \hat{\vartheta}) ds)^2\right) \in \mathbb{R}$$

$$a_2: \mathbb{R} \ni \xi \mapsto a_2(\xi) := \left(\int_0^t \mathbb{E}_\xi(x_s^j) ds\right)_{j \in J_g} \in \mathbb{R}^q$$

and consider again a finite sequence of initial conditions $(\xi_i)_{1 \leq i \leq m}$. Then we also obtain a system of linear equations for the parameters

$$A_2 \vartheta = b_2, \quad (11)$$

with $A_2 := (a_2(\xi_i)^T)_{1 \leq i \leq m} \in \mathbb{R}^{m \times q}$ and $b_2 := (b_2(\xi_i))_{1 \leq i \leq m} \in \mathbb{R}^m$. We define the estimator again via the best approximation

$$\hat{\theta} := A_2^+ b_2. \quad (12)$$

Since the estimation of diffusion parameters θ is based on the estimators $\hat{\vartheta}$ for the drift parameters, additional error sources might affect the estimator $\hat{\theta}$ in practice – see Section 2.3 for an example of this error propagation.

In practice we are confronted with discrete observations instead of continuous ones, so that we need to approximate the (deterministic) integrals in $a_1(\cdot)$, $a_2(\cdot)$, and $b_2(\cdot)$. Assume that we have $(n+1)$ observations at equidistant times $t_k := kh$ for $0 \leq k \leq n$ and $h := t/n$. The goal is to approximate the integrals by means of these observations. Since the integrands depend on the path of the solution of an SDE, we cannot expect the integrands to be very smooth. For such “rough” functions the trapezoidal rule is more accurate than Simpson’s rule [11]. Consequently, we approximate the various integrals via the composite trapezoidal rule

$$\int_0^t \mathbb{E}(x_s^j) ds = \sum_{k=0}^{n-1} \int_{t_k}^{t_{k+1}} \mathbb{E}(x_s^j) ds \approx \frac{h}{2} \left(\mathbb{E}(x_0^j) + \mathbb{E}(x_t^j) + 2 \sum_{k=1}^{n-1} \mathbb{E}(x_{t_k}^j) \right), \quad (13)$$

where we used that $t_0 = 0$ and $t_n = t$.

2.2 Description of the Algorithm: An Example

To apply the actual parametric estimation procedure introduced in the previous section to a specific problem, not only data need to be available but also a parameterization needs to be chosen. Consequently, the complete algorithm can be understood as consisting of two stages:

1. **Initialization:** The time step h is given by the underlying time series of observations and is assumed to be constant (however, this assumption is not necessary, in fact, the procedure might be carried out in the exact same manner with a non-equidistant sampling rate) and the terminal time $t = nh$ is fixed by choosing n appropriately (cf. Section 3). Expectations are approximated by averages over N trajectories generated from N independent Brownian motions. The crucial step is to fix a parameterization for both drift and diffusion coefficients. Lastly, the sequence of initial conditions $(\xi_i)_{1 \leq i \leq m}$ needs to be chosen appropriately.
2. **Two-step Estimation:** Based on the initializations in the previous stage the estimators are well-defined. According to Section 2.1 the parameters ϑ and θ are estimated successively, first the parameters in the drift and then the parameters in diffusion coefficient. For both estimators, two steps need to be performed:
 - a) **Assembling** the linear system equations (9) and (11) respectively.
 - b) **Solving** the arising systems via best approximation (10) and (12) respectively.

Since the estimation step depends on the considered parameterizations for drift and diffusion, we present a detailed pseudocode of the methodology in Algorithm 1 for the example using $J_f = \{1, 3\}$ and $J_g = \{0, 2\}$, i.e. $p = 2 = q$ with

$$f(x; \vartheta) = \vartheta_1 x + \vartheta_3 x^3 \quad \text{and} \quad g(x; \theta) = \theta_0 + \theta_2 x^2. \quad (14)$$

This setting corresponds to the Landau-Stuart equation that will play a vital role in the numerical examples discussed in Section 3. The input arguments of Algorithm 1 are the time step size h and the data array X . The dimension of the array is a result of m different initial conditions each with N trajectories of $(n + 1)$ observations. When we denote by $x_{t|\xi}^i$ the value at time t of the i -th trajectory started initially in ξ , then X corresponds to the collection of these trajectories at discrete (equidistant) times. For the example we consider here, we define approximations of the first three moments at time t via

$$\bar{x}_{t|\xi} := N^{-1} \sum_{i=1}^N x_{t|\xi}^i, \quad \tilde{x}_{t|\xi} := N^{-1} \sum_{i=1}^N (x_{t|\xi}^i)^2, \quad \text{and} \quad \check{x}_{t|\xi} := N^{-1} \sum_{i=1}^N (x_{t|\xi}^i)^3.$$

Thus, the quantities defining the matrices and right-hand sides involved in the estimation step (cf. (9) and (11), respectively) are approximated via

$$a_1(\xi)^T \approx \frac{h}{2} (Q_n(\bar{x}_{\cdot|\xi}), Q_n(\check{x}_{\cdot|\xi}))^T, \quad b_1(\xi) \approx \bar{x}_{t|\xi} - \xi \quad (15a)$$

$$a_2(\xi)^T \approx \frac{h}{2} (2n, Q_n(\tilde{x}_{\cdot|\xi}))^T, \quad b_2(\xi) \approx N^{-1} \sum_{i=1}^N \left(x_{t|\xi}^i - \xi - Q_n(f_{\hat{\vartheta}} \circ x_{\cdot|\xi}^i) \right)^2, \quad (15b)$$

with $f_{\hat{\vartheta}} \equiv f(\cdot; \hat{\vartheta})$ and Q_n denoting the quadrature operator of the trapezoidal rule on $[0, t]$ with n equally spaced ($h = t/n$) subintervals (cf. (13))

$$Q_n(u) := \frac{h}{2} (u_0 + u_t + 2 \sum_{j=1}^{n-1} u_{jh}). \quad (16)$$

Algorithm 1 Algorithm for the estimation of the parameters in the drift and diffusion coefficients in (14).

Require: $h > 0$ and $X \in \mathbb{R}^{m \times N \times (n+1)}$

```

1: for  $i = 1$  to  $m$  do
2:   for  $j = 1$  to  $n$  do
3:      $\alpha_j \leftarrow \sum_{k=1}^N \frac{X_{i,k,j+1}}{N}$ 
4:      $\beta_j \leftarrow \sum_{k=1}^N \frac{(X_{i,k,j+1})^3}{N}$ 
5:   end for
6:    $A_{i,1} \leftarrow \frac{h}{2}(X_{i,1,1} + \alpha_n + 2 \sum_{j=1}^{n-1} \alpha_j)$ 
7:    $A_{i,2} \leftarrow \frac{h}{2}((X_{i,1,1})^3 + \beta_n + 2 \sum_{j=1}^{n-1} \beta_j)$ 
8:    $b_i \leftarrow \alpha_n - X_{i,1,1}$ 
9: end for
10:  $\vartheta \leftarrow A^+ b$ 
11: for  $i = 1$  to  $m$  do
12:   for  $j = 1$  to  $n$  do
13:      $\gamma_j \leftarrow \sum_{k=1}^N \frac{(X_{i,k,j+1})^2}{N}$ 
14:   end for
15:    $A_{i,1} \leftarrow nh$ 
16:    $A_{i,2} \leftarrow \frac{h}{2}((X_{i,1,1})^2 + \gamma_n + 2 \sum_{j=1}^{n-1} \gamma_j)$ 
17:   for  $j = 1$  to  $N$  do
18:      $\delta_j \leftarrow \frac{h}{2}(f(X_{i,1,1}; \vartheta) + f(X_{i,j,n+1}; \vartheta) + 2 \sum_{k=2}^n f(X_{i,j,k}; \vartheta))$ 
19:   end for
20:    $b_i \leftarrow \sum_{j=1}^N \frac{(X_{i,j,n+1} - X_{i,1,1} - \delta_j)^2}{N}$ 
21: end for
22:  $\theta \leftarrow A^+ b$ 
23: return  $(\vartheta^T, \theta^T)^T$ 

```

It should be emphasized that equations such as $x = A^+b$ (e.g. as in lines 10 and 22 in Algorithm 1) are merely meant as a formal notation for x solving the least squares problem $x = \arg \min_{s \in \mathcal{S}} \|s\|_2^2$, $\mathcal{S} = \{z: \|Az - b\|_2^2 \rightarrow \min\}$, cf. equations (10) and (12), rather than indicating that we first compute A^+ and then multiply it with b to obtain x , a step which clearly is computationally inefficient. In practice the method to compute the solution typically depends on the rank of A . For the numerical examples that we will present in Section 3 we ensured a full rank situation by considering a sufficiently large number of different initial conditions. The solution x can be computed with one of the several methodologies for solving least squares problems, e.g. the Cholesky factorization of the normal equations¹ $A^T A x = A^T b$; for details on such methodologies we refer to standard textbooks on numerical linear algebra, e.g. [20, ch. 5].

2.3 Properties of the Estimator

The proposed estimation procedure relies on two key ingredients. The first is that the methodology is based on the identities in (6). The second is that by considering a finite sequence of initial conditions we can cope with multiple parameters in drift and diffusion coefficient, respectively. In the sequel we demonstrate the influence of both components on the proposed estimation scheme with the help of some elementary, nonetheless illustrative, examples when no multiscale effects are present. A detailed and rigorous analysis of the proposed methodology for both multiscale and non-multiscale situations will be presented elsewhere.

To illustrate the influence on the identities in (6) and to address some asymptotic properties we consider a simple Langevin equation with additive noise

$$dx_t = \vartheta f(x_t) dt + \sqrt{\theta} dW_t .$$

The drift estimator proposed in this study – this parameterization is already a straightforward generalization of the one introduced above – relies on the relation

$$\mathbb{E}\left(\int_0^t f(x_s) ds\right)\vartheta = \mathbb{E}(x_t - x_0) , \quad (17)$$

with ϑ being the true value. For a fixed final time $t < \infty$, the estimator for continuous-time observations – meaning that we approximate only the expectation by an average but do not approximate the integrals – based on a single (fixed) initial condition $x_0 = \xi$ is given by

$$\hat{\vartheta} = \frac{\sum_{i=1}^N (x_t^i - x_0)}{\sum_{i=1}^N \int_0^t f(x_s^i) ds} = \vartheta + \frac{\sqrt{\theta} \sum_{i=1}^N \int_0^t dW_s^i}{\sum_{i=1}^N \int_0^t f(x_s^i) ds} = \vartheta + \frac{\mathcal{N}(0, \theta t/N)}{\frac{1}{N} \sum_{i=1}^N \int_0^t f(x_s^i) ds} . \quad (18)$$

Notice that we dropped the dependency on the initial condition, because only a single initial condition is considered here. Since we approximate only the expectations by finite averages, it is not surprising that the estimator for continuous-time observations (18) converges to the true value in agreement with the law of large numbers. The property that the variance of the error vanishes for $N \rightarrow \infty$ reflects the fact that the estimator relies on an identity, i.e. on a direct (deterministic) computation (17) rather than on asymptotic time limits (i.e. on ergodicity).

Recall that we have $(n+1)$ observations at times $0 = t_0 < t_1 < \dots < t_n = t$ with $t > 0$ fixed, in the case of discrete-time observations. The integral approximation introduces an additional error

¹This method for solving the least squares problem is simple to implement but might not be optimal and is usually only recommended for problems with large residuals. However, it is noteworthy that for the numerical examples presented in this study, the method performed well and no stability issues occurred (as can be demonstrated by monitoring the condition number; not shown here), and similar results were obtained with a QR factorization with column pivoting (not shown here).

that can be identified via

$$\int_0^t u_s ds = Q_n(u) + c_n, \quad (19)$$

with an appropriate constant $c_n \in \mathbb{R}$ that depends not only on n but also on u and t , and which vanishes in the limit as $n \rightarrow \infty$. Here Q_n denotes the quadrature operator of the trapezoidal rule as defined in (16). The actual error of the trapezoidal rule depends on the regularity of the integrand. Here we rely only on the assumption $\lim_{n \rightarrow \infty} c_n = 0$ as a general scenario and do not discuss the rate of convergence. Then, similarly to the continuous-time case, the estimator (when neglecting sampling errors) can be written as

$$\begin{aligned} \hat{\vartheta} &= \frac{\sum_{i=1}^N (x_t^i - x_0)}{\sum_{i=1}^N Q_n(f \circ x^i)} = \frac{\sum_{i=1}^N (\vartheta Q_n(f \circ x^i) + \vartheta c_n^i + \sqrt{\theta} \int_0^t dW_s^i)}{\sum_{i=1}^N Q_n(f \circ x^i)} \\ &= \vartheta \left(1 - \frac{\bar{c}_N(n)}{\bar{c}_N(n) - \frac{1}{N} \sum_{i=1}^N \int_0^t f(x_s^i) ds} \right) + \frac{\mathcal{N}(0, \theta t/N)}{\frac{1}{N} \sum_{i=1}^N \int_0^t f(x_s^i) ds - \bar{c}_N(n)}, \end{aligned} \quad (20)$$

with $\bar{c}_N(n) = N^{-1} \sum_{i=1}^N c_n^i$ being the average error constant. Hence, the integral approximation introduces an additional bias – second term in the bracket – that can be controlled by n . Notice that $|\bar{c}_N(n)| \leq \max_{1 \leq i \leq N} |c_n^i|$ so that the additional bias vanishes as $n \rightarrow \infty$ for every N . Consequently, from equation (20) one infers that for a fixed final time $t > 0$, one should choose both $n, N \gg 1$ to obtain an accurate estimate.

To estimate the diffusion coefficient the proposed scheme relies on the relation

$$\theta = \frac{1}{t} \mathbb{E} \left(\left(x_t - x_0 - \vartheta \int_0^t f(x_s) ds \right)^2 \right),$$

that is valid for all times $t > 0$ and where we replace ϑ by its estimator $\hat{\vartheta}$ for concreteness. Consequently, the estimator for continuous-time observation using a single (fixed) initial condition and a fixed final time $t > 0$ reads

$$\begin{aligned} \hat{\theta} &= \frac{1}{tN} \sum_{i=1}^N \left(x_t^i - x_0 - \hat{\vartheta} \int_0^t f(x_s^i) ds \right)^2 = \frac{1}{tN} \sum_{i=1}^N \left((\vartheta - \hat{\vartheta}) \int_0^t f(x_s^i) ds + \sqrt{\theta} \int_0^t dW_s^i \right)^2 \\ &= \theta \frac{1}{N} \chi_N^2 + \frac{(\vartheta - \hat{\vartheta})^2}{tN} \sum_{i=1}^N \left(\int_0^t f(x_s^i) ds \right)^2 + \frac{2(\vartheta - \hat{\vartheta})}{tN} \sum_{i=1}^N W_t^i \int_0^t f(x_s^i) ds. \end{aligned}$$

Since the estimator for the diffusion coefficient depends on the estimated drift parameter, an additional error is introduced (last two terms), as expected. To illustrate the asymptotic properties of the estimator we, however, assume that the error $(\vartheta - \hat{\vartheta})$ is negligible. Consequently, we find that

$$\hat{\theta} \approx \theta \frac{1}{N} \chi_N^2,$$

where χ_N^2 denotes the Chi-squared distribution with N degrees of freedom. Recall that $\frac{1}{N} \chi_N^2 \approx \mathcal{N}(1, 2/N)$ for N sufficiently large, as a consequence of the central limit theorem.

For the modification of discrete-time observations the integrals are again approximated via the trapezoidal rule. Based on the same $(n+1)$ observations (neglecting sampling errors) at $0 = t_0 < t_1 < \dots < t_n = t$ and $t > 0$ fixed we find

$$\hat{\theta} = \frac{1}{tN} \sum_{i=1}^N \left(x_t^i - x_0 - \hat{\vartheta} Q_n(f \circ x^i) \right)^2 = \frac{1}{tN} \sum_{i=1}^N \left((\vartheta - \hat{\vartheta}) \int_0^t f(x_s^i) ds + \sqrt{\theta} \int_0^t dW_s^i + \hat{\vartheta} c_n^i \right)^2,$$

with c_n^i being the error representation of the trapezoidal rule, cf. (19). In contrast to the continuous-time situation, the term $\hat{\vartheta}c_n^i$ reflects the additional error due to the integral approximation that can be controlled by n . Since this additional term is the only difference, expanding the square yields a similar result as in the continuous-time situation. If we assume again that the error from the drift parameter ($\vartheta - \hat{\vartheta}$) is negligible, then we find

$$\hat{\theta} \approx \theta \frac{1}{N} \chi_N^2 + \frac{\hat{\vartheta}^2}{t} \tilde{c}_N(n) + \frac{2\hat{\vartheta}\sqrt{\theta}}{tN} \sum_{i=1}^N c_n^i W_t^i, \quad (21)$$

with $\tilde{c}_N(n) = N^{-1} \sum_{i=1}^N (c_n^i)^2$. Since $0 \leq \tilde{c}_N(n)$ for every N , the second term in equation (21) can only be controlled by n . Consequently, for a fixed final time t , choosing $n, N \gg 1$ it is necessary to obtain an accurate estimate of the true parameter. It is noteworthy that the same steps may be carried out for an arbitrary diffusion function, but obviously we cannot directly infer the distributions of terms involving the stochastic integral.

To deal with multiple parameters in drift and/or diffusion coefficients the proposed estimation scheme relies on considering a finite sequence of initial conditions and defining the estimator via the best approximation. To illustrate the effect of this second component of the estimation procedure, consider the SDE

$$dx_t = \vartheta f(x_t) dt + \sqrt{g(x_t)} dW_t.$$

Provided all quantities are well defined, the estimator of ϑ , via a best approximation using a sequence of initial conditions $(\xi_i)_{1 \leq i \leq m}$, is given by

$$\hat{\vartheta} = \frac{\sum_{i=1}^m a_1(\xi_i) b_1(\xi_i)}{\sum_{i=1}^m a_1(\xi_i)^2},$$

where $a_1(\cdot)$ and $b_1(\cdot)$ are as in (8) associated with the drift function f . On the other hand, the quantity $b_1(\xi_i)/a_1(\xi_i)$ yields a local estimator for each initial condition because the considered problem has only one parameter to be determined. Thus the best approximation corresponds here to the weighted arithmetic mean of these local estimators with weights $a(\xi_i)^2$. Consequently, increasing m includes an additional stabilization effect into the estimation scheme. From this point of view, the best approximation resolves naturally the problem of combining local estimates to a global estimator that arises in different estimation procedures as well, where piecewise local strategies are utilized to improve the estimates; see for example [7]. In this study a heuristic estimation strategy for non-constant diffusions in the MLE framework is proposed by extracting local information from a large time-series to estimate coefficients locally and combine these local estimators to global estimators. Since this strategy is based on the MLE, the results in the above study also highlight the subsampling issue when applied to multiscale diffusions.

It should be noted that although the proposed methodology computes moments of the solution of an SDE, there is no direct link to the generalized method of moments (GMM) [23]. The GMM for parametric estimation for SDEs is more closely related to the MLE instead, as both estimation schemes rely on ergodicity of the corresponding process by using long time-series. In fact, both estimators even coincide for many cases [22, ch. 14.4]. Since the MLE is biased for multiscale diffusions, it can be expected that the subsampling issue also arises for the GMM when applied to data obtained from a system with multiscale structure.

2.4 Estimators for Multiscale Diffusions

In the context of diffusion processes with two widely separated time scales we consider the following set-up: A fast/slow system of SDEs

$$dx_t = \left(\frac{1}{\varepsilon} f_1(x_t, y_t) + f_0(x_t, y_t) \right) dt + \alpha(x_t, y_t) dU_t, \quad (22a)$$

$$dy_t = \left(\frac{1}{\varepsilon^2} g_2(x_t, y_t) + \frac{1}{\varepsilon} g_1(x_t, y_t) + g_0(x_t, y_t) \right) dt + \frac{1}{\varepsilon} \beta(x_t, y_t) dV_t, \quad (22b)$$

equipped with appropriate initial conditions. In (22) U, V denote Brownian motions of appropriate dimensions and $0 < \varepsilon \ll 1$ denotes a small parameter. For the dimension of the fast/slow system we assume that $y: T \mapsto \mathbb{R}^d$ and (for simplicity) $x: T \mapsto \mathbb{R}$, where $T = [0, t]$ denotes a finite time interval of interest. Furthermore, we assume that drift and diffusion functions in (22a) and (22b) respectively, are such that there exists a well-defined (i.e. the SDE provides a unique weak solution on any finite time interval and for any initial condition) coarse-grained SDE

$$dX_t = f(X_t) dt + \sqrt{g(X_t)} dW_t, \quad (23)$$

in the limit of $\varepsilon \rightarrow 0$; see e.g. [47, ch. 11 and 18] and references therein for technical details. That is, the slow process x is approximated by the solution of (23) for $\varepsilon \ll 1$. Even in cases where both the effective drift f and the effective diffusion function g are known, the actual computation of these expressions might be difficult or even impossible. Hence, our goal is to estimate both effective drift coefficient f and effective diffusion coefficient g in (23) from available data (observations) of the fast/slow system (22) (more precisely, only of its slow component). To this end we assume the same parameterizations of the effective drift and diffusion functions as introduced in Section 2.1:

$$f(x) \equiv f(x; \vartheta) := \sum_{j \in J_f} \vartheta_j x^j \quad \text{and} \quad g(x) \equiv g(x; \theta) := \sum_{j \in J_g} \theta_j x^j,$$

where we recall that $J_f, J_g \subset \mathbb{N}_0$ denote index sets with $p = |J_f|$ and $q = |J_g|$. Our goal then is: Given observations of the slow component (22a), is it possible to estimate the parameters $\vartheta \equiv (\vartheta_j)_{j \in J_f} \in \mathbb{R}^p$ and $\theta \equiv (\theta_j)_{j \in J_g} \in \mathbb{R}^q$? Under the assumption that the (ergodic) fast process (22b) is stationary, the algorithm described in Section 2.1 applies straightforwardly also for this problem, given the final time t of observation length is appropriately chosen; cf. Section 3.2. That is, using multiple initial conditions for the slow process (22b) to deal with multiple parameters in drift and diffusion, while the fast process is sampled from its invariant measure that is assumed to be known, either analytically or numerically.

The main motivation to consider the same algorithm also in the presence of multiscale effects originates from the fact that both the slow component of the full fast/slow system and the effective dynamics have probability laws that are approximately the same, provided $\varepsilon \ll 1$. Consequently, also expectations with respect to these laws are approximately equal. Since the proposed methodology is based on expectations, cf. equation (6), we believe that this approach yields asymptotically unbiased estimators. A rigorous analysis of the proposed methodology to verify this heuristic argument is currently work in progress and will be reported elsewhere.

We conclude this section with a remark on a recently proposed estimator for constant diffusion coefficients [16] that can also be derived using the approach we introduce here. To this end, assume that the coarse-grained equation takes the form:

$$dx_t = f(x_t) dt + \sqrt{\theta} dW_t,$$

where we assume that the effective drift f is known and we wish to estimate the diffusion coefficient θ from available data of a fast/slow system. When considering a single (fixed) initial condition and

following the approach introduced in Section 2.1, the resulting estimator reads

$$\hat{\theta} = \frac{1}{tN} \sum_{i=1}^N \left((x_t^i - x_0 - \sum_{j=0}^{n-1} \int_{jh}^{(j+1)h} f(x_s^i) ds)^2 \right),$$

where the integrals are being approximated by a quadrature rule and t is chosen appropriately (in fact, we have approximated the Lebesgue integral via the trapezoidal rule). If one, however, uses a rectangular-method with the left corner node instead, this estimator coincides with the estimator proposed in [16] for estimating the effective diffusion coefficient based on observations of the slow component of a fast/slow system. We emphasize, that a crucial assumption on the estimator in the aforementioned work is that the effective drift is known a priori. This assumption is very restrictive and makes the estimator unfeasible for most practical applications. A further limitation of the estimator in [16] is that it applies only in situations where the noise in the coarse-grained equation is additive (constant diffusion coefficient). Conversely, the methodology proposed here aims to estimate multiple parameters in both the drift and the diffusion coefficients.

3 Numerical Experiments

We now present numerical experiments of parameter estimations for diffusion processes to illustrate the behavior of the estimation scheme developed in Section 2.

3.1 Parameter Estimation for single-scale SDEs

Let us first present numerical results for parameter estimation when no multiscale effects are present. We investigate two different SDEs. In Section 3.1.1 we consider the SDE corresponding to the Ornstein-Uhlenbeck process and in Section 3.1.2 we examine the stochastic Landau-Stuart equation. The purpose of these examples is twofold. On the one hand they are used to illustrate that the proposed methodology can be employed successfully in practice to estimate unknown parameters and on the other they help us understand the influence of the parameters n , h , N , and m on the algorithmic estimation procedure. The time series were obtained by solving the corresponding SDEs via the Euler-Maruyama scheme. In all numerical examples reported in this section, we used a time step of $h = 10^{-3}$.

3.1.1 Ornstein-Uhlenbeck Process

Consider the following SDE

$$dx_t = -Ax_t dt + \sqrt{\sigma} dW_t, \quad x_0 = \xi, \quad (24)$$

with the unique solution being the Ornstein-Uhlenbeck process starting at ξ . The estimation procedure is applied to data from the SDE with true parameters $(A, \sigma) = (0.5, 0.5)$ using a variety of different values for the parameters of the algorithm. Figure 1 depicts the relative errors of the estimated values as functions of the number of initial conditions m for different combinations of n (recall that the final time is $t = nh$) and N . For the estimated drift parameter \hat{A} (Figure 1(a)) there is a discrepancy among different combinations of n and N for small values of m and the relative errors vary from approximately 0.08 to 6. Increasing m decreases the relative errors, with fluctuations due to the discretization, as expected; cf. Section 2.3. Moreover, it is apparent that the larger n and N , the smaller the relative error. In contrast to the drift parameter, the relative errors of the diffusion parameter $\hat{\sigma}$ (Figure 1(b)) are already small (relative error smaller than 0.05) even for small values of m . This is due to the constant diffusion coefficient of the SDE. Increasing m generally decreases the relative error further, again, with fluctuations due to the discretization.

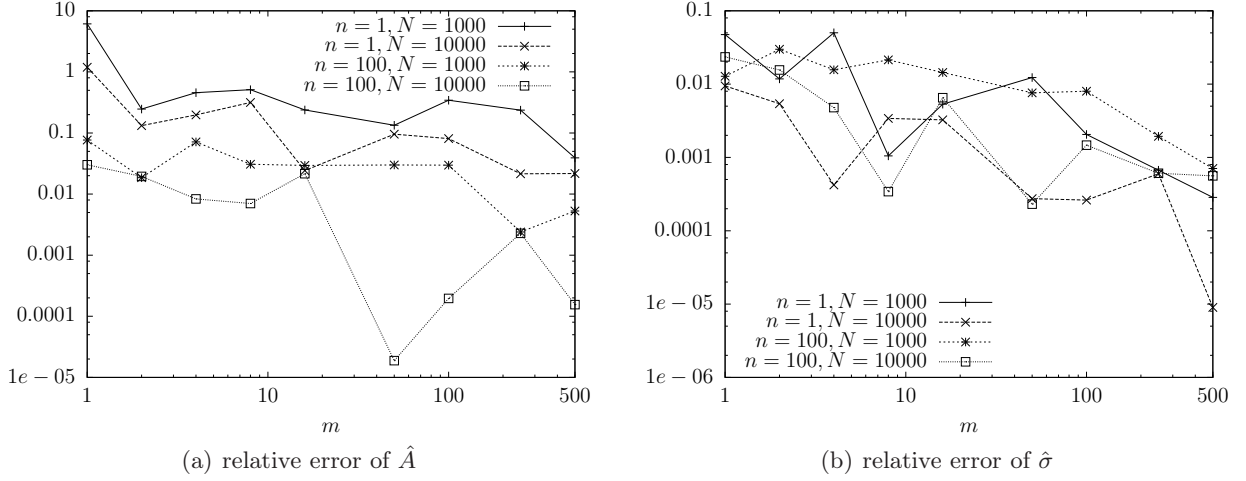


Figure 1: Relative error of the estimated parameters in (24) as functions of the number of initial conditions m in a log-log scale. The final time of the considered time series is $t = nh$ with $h = 0.001$ and the true parameters are $(A, \sigma) = (0.5, 0.5)$. Furthermore, N denotes the number of independent Brownian paths.

We note that although an error propagates from the drift estimation to the diffusion estimation (cf. Section 2), it appears to be negligible in this example. Based on these results, it seems plausible to tune the algorithm-defining parameters in a way such that the estimators provide a given relative accuracy while at the same time the computational cost is minimized. Obviously the question of optimized algorithm-defining parameters is of high importance in practical applications. However, this is not a straightforward issue to address as not only the discretization related parameters n and N influence the accuracy of the algorithm in practice, but also the number of initial conditions m as well as their locations; we will investigate these and related issues in a future study, see also the discussion in Section 4.

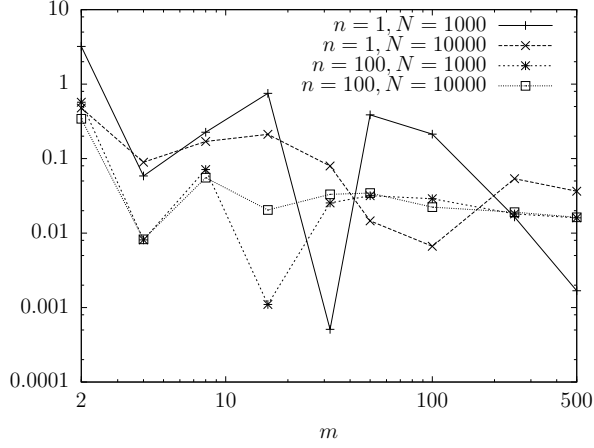
3.1.2 Landau-Stuart Equation

Consider the stochastic Landau-Stuart equation [35, ch. 2.2], where both additive and multiplicative noise are present

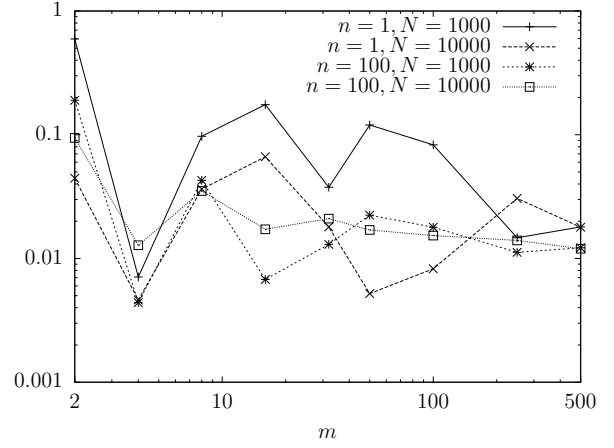
$$dx_t = (Ax_t - Bx_t^3) dt + \sqrt{\sigma_a + \sigma_b x_t^2} dW_t, \quad x_0 = \xi. \quad (25)$$

This SDE can be obtained from a wide class of spatially extended systems, e.g. the noisy Kuramoto-Sivashinsky equation [50, 49] by assuming near-critical conditions, i.e. being sufficiently close to the primary bifurcation, and employing the homogenization theory developed in [6]. In this case we need to estimate a total number of four parameters, two in the drift and two in the diffusion coefficient.

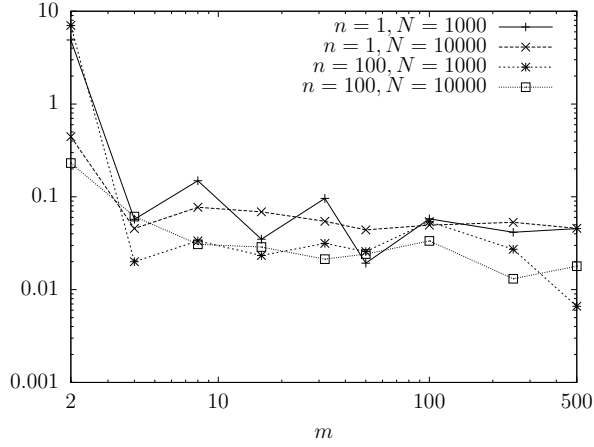
We performed various numerical experiments with different choices of parameters. Figure 2 illustrates the relative error of the estimated parameters as functions of the number of initial conditions for different combinations of n and N when the true parameters are $(A, B, \sigma_a, \sigma_b) = (3, 2, 1.5, 1.3)$. We find qualitatively the same behavior as in the previous section for the Ornstein-Uhlenbeck process: all three parameters n , N , and m affect the accuracy of the estimators. Although the Figures 2(a)–(d) show a decreasing trend of the relative errors when increasing m , fluctuations are still present. These fluctuations are of different magnitude for different parameters and are reduced by increasing both n and N . One also observes that increasing m improves the accuracy of the estimators but only up to a certain level, i.e. the corresponding curves approach nearly constant values for large m , that are due to the approximation of the original system matrix, for instance



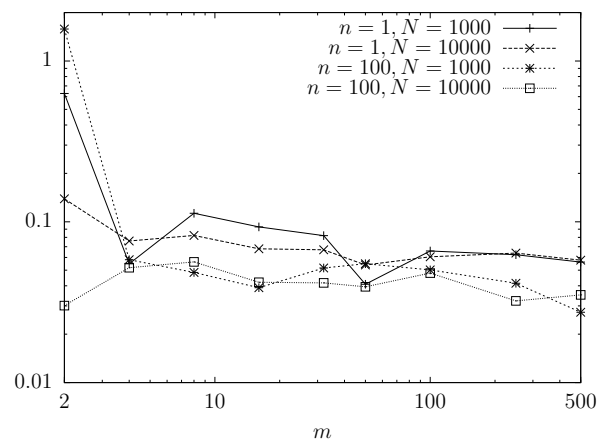
(a) relative error of \hat{A}



(b) relative error of \hat{B}



(c) relative error of $\hat{\sigma}_a$



(d) relative error of $\hat{\sigma}_b$

Figure 2: Relative error of the estimated parameters in (25) as functions of the number of initial conditions m using a log-log scale. The final time of the considered time series is $t = nh$ with $h = 0.001$ and the true parameters are $(A, B, \sigma_a, \sigma_b) = (3, 2, 1.5, 1.3)$. N denotes the number of independent Brownian paths.

in (9), by a matrix based on observations and the aforementioned discretizations. Although it is apparent that estimating all parameters in the Landau-Stuart model is more delicate than for the Ornstein-Uhlenbeck process in Section 3.1.1, it nonetheless seems possible to determine optimal algorithm-defining parameters such that the computational cost is minimized given a certain error tolerance for the estimators for this model also.

3.2 Parameter Estimation for Fast-Slow Systems

Of particular interest is the behavior of the estimator when applied to systems with two different time scales. We examine the properties of the estimation scheme for stochastic multiscale diffusions (Sections 3.2.1–3.2.3), truncated systems of time rescaled stochastic partial differential equations (Section 3.2.4), and deterministic systems that exhibit temporal chaos that can be approximated by an appropriate SDE (Section 3.2.5). As the precise dependency of the estimation procedure on the control parameters n, m, N, h is still an open question, the main purpose of this section is to illustrate the general applicability of the proposed estimation procedure for multiscale diffusions. If not stated otherwise, the generated time series were obtained by solving the corresponding multiscale SDEs via the Euler-Maruyama scheme using a time step $h = 10^{-3}$. Furthermore, the expectation is approximated by an average using $N = 5000$ independent Brownian paths and $m = 150$ different (equally-spaced) initial conditions are used. We emphasize once again that this particular choice of algorithm-defining parameters might be far from optimal in the sense of computational complexity. But since our main goal is to demonstrate the applicability of the proposed scheme to multiscale diffusions, these algorithm-defining parameters will yield reliable estimators.

3.2.1 Fast Ornstein-Uhlenbeck Noise

When the fast process is an Ornstein-Uhlenbeck process it is rather straightforward to determine the precise form of the effective equation associated to the fast/slow system, because this task reduces to computing Gaussian integrals. Consider for example

$$dx_t = \left(\frac{1}{\varepsilon} \sigma(x_t) y_t + h(x_t, y_t) - \sigma'(x_t) \sigma(x_t) \right) dt, \quad (26a)$$

$$dy_t = -\frac{1}{\varepsilon^2} y_t dt + \frac{\sqrt{2}}{\varepsilon} dV_t, \quad (26b)$$

with V_t being a standard Brownian motion, then the effective dynamics is given by

$$dX_t = \bar{h}(X_t) dt + \sqrt{2\sigma(X_t)^2} dW_t, \quad (27)$$

where $\bar{h}(x)$ denotes the average of $h(x, \cdot)$ with respect to the invariant measure of the fast process (Ornstein-Uhlenbeck process). We note that we have subtracted the Stratonovich correction from the drift in (26a), so that the noise in (27) can be interpreted in the Itô sense.² In the sequel we consider two different choices of the pair $h(\cdot), \sigma(\cdot)$. As a first example let

$$h(x, y) = h(x) = Ax \quad \text{and} \quad \sigma(x) = \sqrt{\sigma}, \quad (28)$$

then the amplitude equation is precisely the Ornstein-Uhlenbeck process (cf. Section 3.1.1) but here in the context of multiscale diffusions where classical estimators fail. To illustrate this failure and motivate the necessity of an appropriate sampling rate, the quadratic variation of the path (QVP)

²The noise entering (26a), i.e. the process (26b), is a smoothed approximation to white noise, so that the noise in the limiting equation has to be interpreted in the Stratonovich sense according to the Wong-Zakai theorem. Correcting the drift is not essential for the applicability of our methodology, it was done so that the limiting equation (27) is somewhat simpler.

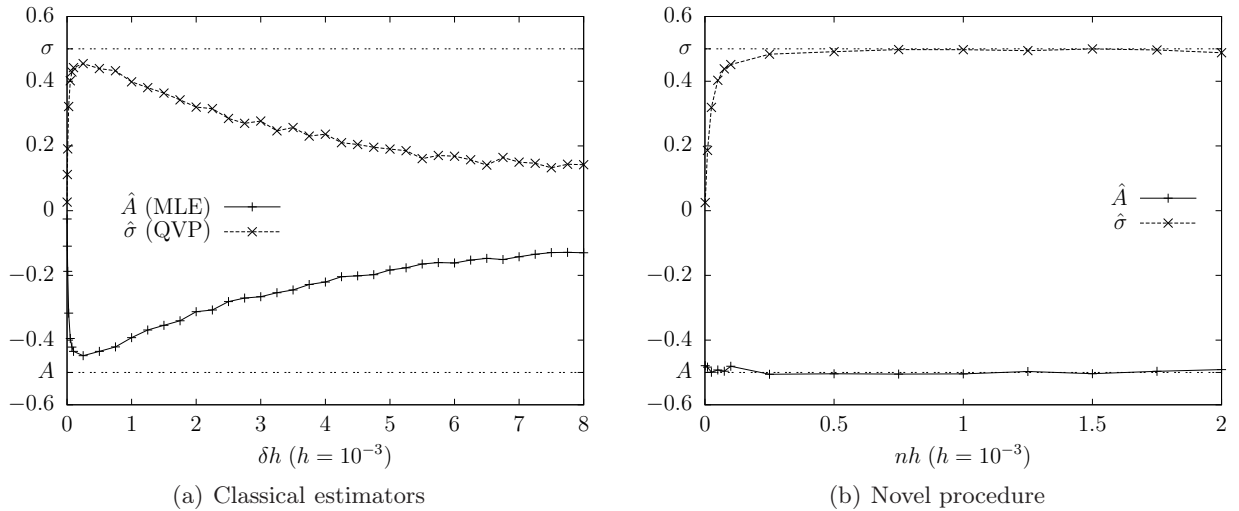


Figure 3: Performance of classical estimators and our procedure for both drift and diffusion coefficients in (28). In Figure (a) the MLE and QVP as functions of the subsampling rate δh ($\delta = 1$ corresponds to no subsampling) are shown. Figure (b) shows the result of our procedure as a function of the final time $t = nh$. In both cases the sampling rate of the considered time series is $h = 0.001$ and the true parameters are $(A, \sigma) = (-0.5, 0.5)$. Dashed lines denote the true values.

estimator for the effective diffusion constant and the maximum likelihood estimator (MLE) for the effective drift parameter are applied to a time series on $[0, 5000]$ with initial condition $x_0 = 0.5$ generated by the associated fast/slow system (26) with true parameters $(A, \sigma) = (-0.5, 0.5)$ and $\varepsilon = 0.1$. Figure 3(a) depicts the performance of both the QVP and the MLE as functions of the subsampling rate δh . The parameter δ indicates here that only every δ -th observation is used to estimate the parameters in the drift and diffusion coefficients, hence δh denotes the time period between two consecutive observations. Starting from the case without subsampling ($\delta = 1$, that is $\delta h = 10^{-3}$) and increasing δ , both estimators approach the true value. However, after an optimal subsampling rate, for which the estimator is as close to the true value as possible (here approximately $\delta h = 0.25$, that is $\delta = 250$), both estimators deviate monotonically from the target value. We note that at the optimal subsampling rate the relative error is approximately 10%. Figure 3(a) seems to suggest that the optimal subsampling rate is given by the local extremum of the estimator as function of the subsampling. However, this behavior is not true in general, see for instance the numerical examples in [46, 16] that reveal different behaviors. Recall, that the optimal subsampling rate is in general unknown.

The situation is different when the effective parameters are estimated via the method introduced here. Observations are generated by the associated fast/slow system (26) with true parameters $(A, \sigma) = (-0.5, 0.5)$ and $\varepsilon = 0.1$. The performance of the estimator as function of the final time $t = nh$ for both \hat{A} and $\hat{\sigma}$ is plotted in Figure 3(b) and is compared directly to the results obtained by the classical estimators discussed before. For small values of $t = nh$ one observes that the estimated value of the drift parameter \hat{A} fluctuates around the true value and stabilizes for larger times with minor fluctuations around the true value. We note that the estimator obtained by the proposed scheme significantly outperforms the MLE in terms of accuracy.³ For the estimated diffusion coefficient

³At the optimal subsampling rate the relative error for the MLE is approximately 10%, whereas the relative error for the drift parameter obtained by the novel scheme is less than 1% for $t = nh \geq 0.25$. Admittedly, this comparison is not completely fair, because the novel estimation scheme in its current form relies on more data than the MLE, but we believe that these numbers nonetheless illustrate the potential of the methodology we propose.

(a) Classical Setting			(b) Multiscale Setting		
N	$\hat{\sigma}_a$	$\hat{\sigma}_b$	N	$\hat{\sigma}_a$	$\hat{\sigma}_b$
1	0.948301	0.251906	1	0.056463	0.000000
10	0.871561	0.390639	10	0.039087	0.028682
100	0.817719	0.480469	100	0.040613	0.026214
1000	0.806024	0.500243	1000	0.040363	0.026727

Table 1: QVP estimators when fitting the Landau-Stuart SDE (30) to observed data. For the results outlined in table (a) the observed data were obtained from (30), i.e. the classical setting, whereas for the results in table (b) observations of the slow component (29a) of the fast/slow system (29) were used, i.e. the multiscale setting. In both cases the true parameters are $(\sigma_a, \sigma_b) = (0.81, 0.49)$ and $\varepsilon = 0.1$ was used in the multiscale setting. The parameter N indicates the number of independent Brownian paths that have been used to compute an average.

cient $\hat{\sigma}$ one finds that the novel scheme proposed here and the QVP estimator yield similar results for small values of $t = nh$, indicating that increasing $t = nh$ here has the same beneficial effect as subsampling does for the QVP estimator. But unlike the QVP, the estimator proposed here approaches the correct value further and closely fluctuates around it when increasing nh , without any deviations as when using QVP and subsampling. This is a typical behavior for the estimator when applied to multiscale diffusions as we will see in the forthcoming examples. Consequently, one finds that, unlike classical estimators, once the final time $t = nh$ is larger than a critical value, the estimator fluctuates closely around the true value.

Consider as a second example $h(x, y) = h(x) = Ax - Bx^3$ and $\sigma(x) = \sqrt{\sigma_a + \sigma_b x^2}$, so that the fast/slow system (26) reads

$$dx_t = \left(\frac{1}{\varepsilon} y_t \sqrt{\sigma_a + \sigma_b x_t^2} + (A - \sigma_b) x_t - B x_t^3 \right) dt, \quad (29a)$$

$$dy_t = -\frac{1}{\varepsilon^2} y_t dt + \frac{\sqrt{2}}{\varepsilon} dV_t, \quad (29b)$$

with the effective dynamics in (27) given by the Landau Stuart equation (see Section 3.1.2)

$$dX_t = (AX_t - BX_t^3) dt + \sqrt{2(\sigma_a + \sigma_b X_t^2)} dW_t. \quad (30)$$

A natural extension of the QVP estimator to diffusion coefficients that depend on multiple parameters is obtained by considering the standard QVP relation for different time increments. Provided that one considers a sufficient number of increments, i.e. a sufficient number of estimating equations, one can define the QVP estimator by solving the arising system. To illustrate that this approach can deal with multiple parameters in the diffusion coefficient when no multiscale effects are present in the data, we first use the QVP to estimate σ_a and σ_b in (30) based on data that are also obtained from (30). This (single-scale) situation corresponds to the classical case of parametric estimation and the QVP can indeed be used to obtain accurate estimators $\hat{\sigma}_a$ and $\hat{\sigma}_b$, as it is illustrated in table 1(a). Therein the obtained estimators $\hat{\sigma}_a$ and $\hat{\sigma}_b$ are presented for different values of N , where N indicates the number of independent Brownian paths that have been used to compute an average to improve the accuracy. The Landau-Stuart equation (30) was solved numerically on $T = [0, 1000]$ starting at $X(0) = 0.5$ with true parameters $(A, B, \sigma_a, \sigma_b) = (1, 2, 0.81, 0.49)$. Apparently, the QVP yields very accurate estimates in this setting. However, things change when the QVP is adopted in

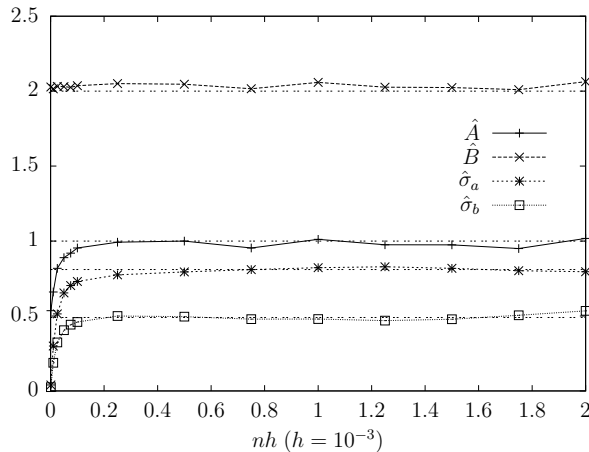


Figure 4: Performance of the novel estimators \hat{A} , \hat{B} , $\hat{\sigma}_a$, and $\hat{\sigma}_b$ for the Landau-Stuart equation (30) as functions of the final time $t = nh$ with $h = 0.001$. The true effective parameters are $(A, B, \sigma_a, \sigma_b) = (1, 2, 0.81, 0.49)$.

the presence of multiple time scales. In this multiscale setting we wish to estimate the parameters σ_a and σ_b in the effective dynamics (30) from observations of the slow component (29a) of the fast/slow system (29). The same true parameters and configuration (i.e. same number and length of time increments) of the QVP as in the classical example without multiscale effects were used, since the QVP performed well therein. Table 1(b) displays the obtained estimators $\hat{\sigma}_a$ and $\hat{\sigma}_b$ for different values of N when the observations are obtained from (29) with scale separation $\varepsilon = 0.1$. Increasing N yields QVP estimators with minor fluctuations but both estimators are strongly biased, as expected. Hence, an appropriate subsampling of the data would be required to remove the bias, but, once again the optimal subsampling rate is not known a priori and might, as in the case of the MLE, even be different for different parameters.

Conversely, we will use this example to illustrate that the parameters in multiscale diffusions can be estimated accurately using the proposed scheme, even though the amplitude equation provides a far more involved structure than the one of the previous example in addition to the multiscale structure of the problem. Figure 4 depicts the performance of the estimation procedure based on observations generated by the fast/slow system (29) with true parameters $(A, B, \sigma_a, \sigma_b) = (1, 2, 0.81, 0.49)$ and $\varepsilon = 0.1$ as a function of the final time $t = nh$. The true values are indicated by dashed lines. The behavior of the estimators is qualitatively similar with that in the previous example. By increasing $t = nh$, the estimators approach the true values, respectively and fluctuate closely around them after a critical final time. Consequently, all parameters can be estimated accurately.

3.2.2 Brownian motion in a two-scale Potential: A Quadratic Potential in one dimension

Here we study the example that was originally used in [46] to illustrate the failure of classical estimation schemes in the context of multiscale diffusions for the first time. More precisely we consider the first-order Langevin equation

$$dx_t = -\nabla V_\alpha\left(x_t, \frac{x_t}{\varepsilon}\right) dt + \sqrt{2\sigma} dU_t, \quad (31)$$

which is a simple model to describe the movement of a Brownian particle in a two-scale potential V_α subject to thermal noise – U_t being a standard Brownian motion. Here we consider the one-dimensional problem (a two-dimensional example is treated in see Section 3.2.2) and further assume that the two-scale potential is given by a large scale as well as a fluctuating part: $V_\alpha(x, y) =$

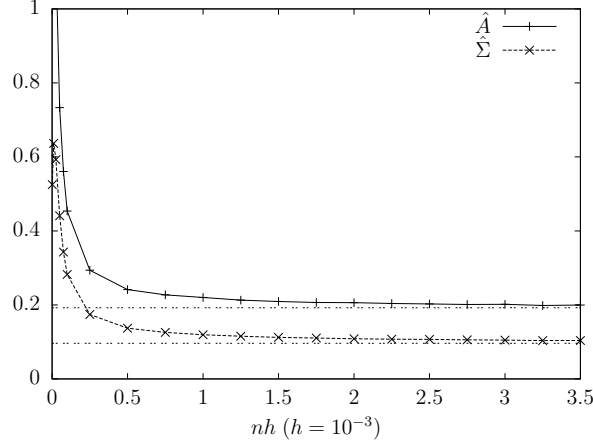


Figure 5: Performance of the estimators $\hat{A}, \hat{\Sigma}$ in (32) as functions of the final time $t = nh$ with $h = 0.001$.

$\alpha V(x) + p(y)$. Based on these assumptions we can rewrite the Langevin equation as

$$dx_t = -\left(\alpha V'(x_t) + \frac{1}{\varepsilon} p'\left(\frac{x_t}{\varepsilon}\right)\right) dt + \sqrt{2\sigma} dU_t.$$

When the fluctuating part p is sufficiently smooth and periodic with period L , the effective dynamics is given by

$$dX_t = -AV'(X_t) dt + \sqrt{2\Sigma} dW_t, \quad (32)$$

where the effective coefficients are given by $A = \alpha L^2 / (Z_+ Z_-)$ and $\Sigma = \sigma L^2 (Z_+ Z_-)$, where $Z_{\pm} = \int_0^L e^{\pm p(y)/\sigma} dy$, see [46] for details. Here we consider $V(x) = x^2/2$ and $p(y) = \cos(y)$, so that (32) is the SDE of an Ornstein-Uhlenbeck process with

$$A = \frac{\alpha}{I_0(\sigma^{-1})^2} \quad \text{and} \quad \Sigma = \frac{\sigma}{I_0(\sigma^{-1})^2},$$

where $I_0(z)$ denotes the modified Bessel function of first kind, cf. [2, ch. 9.6]. We note that both the effective drift and the effective diffusion depend on the diffusion σ of the original fast/slow system⁴. Figure 5 shows the performance of the estimation scheme when applied to observations of the fast/slow system with $(\alpha, \sigma) = (1, 0.5)$ and $\varepsilon = 0.1$. As for the examples in the previous section both estimators \hat{A} and $\hat{\Sigma}$ are biased for small final times $t = nh$. Using longer time series, i.e. increasing nh , reduces this bias and both estimators approach the true values (dashed lines) respectively.

3.2.3 Brownian motion in a two-scale potential: A quadratic potential in two dimensions

To illustrate that the proposed methodology can readily be extended to multivariate processes, we consider here a generalization of (31) in two dimensions

$$dx_t = -\nabla V\left(x_t, \frac{x_t}{\varepsilon}; M\right) dt + \sqrt{2\sigma} dU_t,$$

⁴We remark that in the numerical examples presented in [16, sec. 4.3] the drift coefficient in the homogenized equation is assumed to be known when estimating the diffusion coefficient, even though it depends explicitly on the unknown σ .

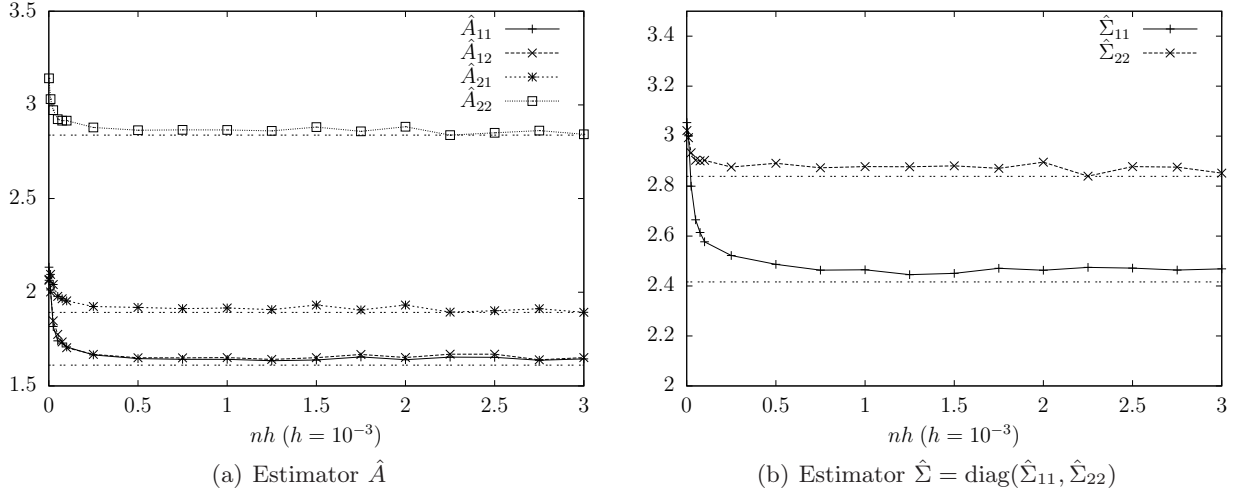


Figure 6: Performance of the estimators $\hat{A}, \hat{\Sigma}$ in (33) as functions of the final time $t = nh$ with $h = 0.001$.

where $V(\cdot, \cdot; M)$ denotes again a two-scale potential with M being a set of parameters controlling the drift and U_t denotes a standard two-dimensional Brownian motion. As with the one-dimensional case, we assume that the two-scale potential $V(\cdot, \cdot; M)$ is given by a large scale as well as a fluctuating part, with the fluctuating part being separable: $V(x, y; M) = V(x; M) + p_1(u) + p_2(v)$, with $x, y \in \mathbb{R}^2$ and $y = (u, v)^T$. Hence, the original system reads

$$dx_t = -\left(\nabla V(x_t; M) + \frac{1}{\varepsilon} \begin{pmatrix} p'_1(x_t^1/\varepsilon) \\ p'_2(x_t^2/\varepsilon) \end{pmatrix}\right) dt + \sqrt{2\sigma} dU_t,$$

with $x_t = (x_t^1, x_t^2)^T \in \mathbb{R}^2$. We take the large scale part to be a quadratic potential

$$V(x; M) = \frac{1}{2} x^T M x,$$

with M being symmetric and positive definite, so that the effective dynamics is given by

$$dX_t = -KM X_t dt + \sqrt{2\sigma K} dW_t, \quad (33)$$

for $X_t \in \mathbb{R}^2$ with analytic expressions for $K = \text{diag}(k_1, k_2)$; see [46]. With $p_1(u) = \cos(u)$ and $p_2(v) = \cos(v)/2$ we find

$$k_1 = \frac{1}{I_0(1/\sigma)^2} \quad \text{and} \quad k_2 = \frac{1}{I_0(1/(2\sigma))^2},$$

where $I_0(z)$ denotes again the modified Bessel function of the first kind.

Since both identities in (6) have pendants for multivariate processes, the methodology introduced here can be readily applied to estimate both effective drift matrix $A := KM$ and the effective diffusion matrix $\Sigma := 2\sigma K$ in (33). The only difference is that the system of equations corresponding to (9) and (11), respectively, is a matrix equation in this case, and similar techniques to obtain the best approximation (both formally and numerically) can be employed [48]. Figure 6 depicts the performance of the estimation scheme when applied to observations of the fast/slow system with $M = \begin{pmatrix} 2 & 2 \\ 2 & 3 \end{pmatrix}$, $\sigma = 3/2$, and $\varepsilon = 0.1$. Figure 6(a) shows the estimated values of the four effective drift coefficients: As for the one-dimensional examples, the estimators also show here an approaching behavior towards the target values when increasing $t = nh$, yielding an accurate

estimate of $A = KM$ for $t = nh$ sufficiently large. The same behavior when increasing $t = nh$ is also observed for the estimated (diagonal) effective diffusion coefficient $\Sigma = 2\sigma K$, as it is shown in Figure 6(b). We note that although the curves in Figure 6(b) show a minor gap for larger $t = nh$, the relative error is less than 2% in both cases for $t \geq 0.75$.

3.2.4 Truncated Burgers' Equation

We consider here an appropriately rescaled variant of the stochastic Burgers' equation

$$du_t = \left(\frac{1}{\varepsilon^2}(\partial_x^2 + 1)u_t + \frac{1}{2\varepsilon}\partial_x u_t^2 + \nu u_t \right) dt + \frac{1}{\varepsilon}\mathcal{Q}d\mathcal{W}_t$$

on $[0, \pi]$ equipped with homogeneous Dirichlet boundary conditions. Under technical assumptions on the covariance operator \mathcal{Q} of the space-time white noise \mathcal{W} , one can show (see [1] and references therein for details) that the coefficients of the three-term truncated representation of the solution have to solve the following multiscale SDE

$$\begin{aligned} dx_t &= \left(\nu x_t - \frac{1}{2\varepsilon}(x_t y_t^1 + y_t^1 y_t^2) \right) dt, \\ dy_t^1 &= \left(\nu y_t^1 - \frac{3}{\varepsilon^2}y_t^1 - \frac{1}{2\varepsilon}(2x_t y_t^2 - x_t^2) \right) dt + \frac{q_1}{\varepsilon} dV_t^1, \\ dy_t^2 &= \left(\nu y_t^2 - \frac{8}{\varepsilon^2}y_t^2 + \frac{3}{2\varepsilon}x_t y_t^1 \right) dt + \frac{q_2}{\varepsilon} dV_t^2, \end{aligned}$$

with V_t^1 and V_t^2 being independent standard Brownian motions. Notice that noise acts only on the fast modes directly. For the truncated system standard homogenization theory applies and yields

$$dX_t = (AX_t - BX_t^3) dt + \sqrt{\sigma_a + \sigma_b X_t^2} dW_t \quad (34)$$

as the effective dynamics with true parameters

$$A = \nu + \frac{q_1^2}{396} + \frac{q_2^2}{352}, \quad B = \frac{1}{12}, \quad \sigma_a = \frac{q_1^2 q_2^2}{2112}, \quad \text{and} \quad \sigma_b = \frac{q_1^2}{36}.$$

See [6] for details. Figure 7 shows the performance of the estimation scheme when applied to observations of the three dimensional fast/slow system with $\nu = 1$, $(q_1, q_2) = (1, 1)$, and $\varepsilon = 0.1$. Since the true values of the effective coefficients (dashed lines) are of different orders for these particular choices, a semi-logarithmic scale is adopted for the sake of clarity. The plots show qualitatively the same behavior as in the previous examples when increasing the final time $t = nh$ and suggest that the estimation procedure yields accurate estimators. Only the estimated value $\hat{\sigma}_a$ fluctuates around the true value. We note that the true value σ_a is very small ($\approx 5 \cdot 10^{-4}$) so that this coefficient has only marginal influence in the complete diffusion function. Furthermore, recall that even the time-step in the Euler-Maruyama discretization is larger ($h = 10^{-3}$), so that the fluctuations are expected to be due to discretization errors. As a matter of fact, considering a finer discretization (i.e. increasing m, n , and N) reduces the fluctuations (not shown here).

3.2.5 Fast chaotic noise

Our methodology is also applicable to a system of ordinary differential equations (ODEs) where the stochastic noise is replaced by deterministic chaos. In particular, we will consider an ODE driven by one of the components of an appropriately rescaled Lorentz system. More precisely, consider as

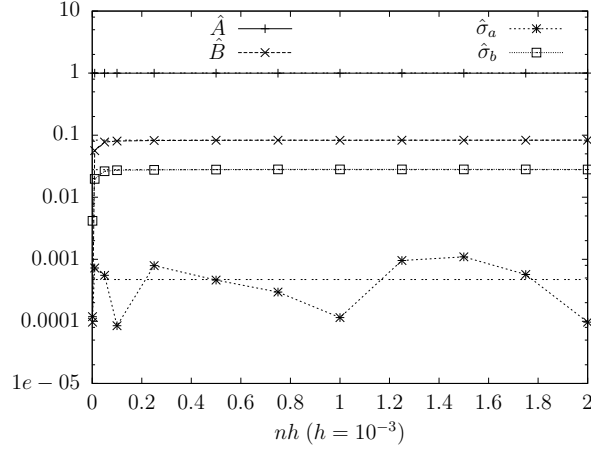


Figure 7: Performance of the estimators \hat{A} , \hat{B} , $\hat{\sigma}_a$, and $\hat{\sigma}_b$ in (34) as functions of the final time $t = nh$ with $h = 0.001$.

an example the following system

$$\frac{dx}{dt} = x - x^3 + \frac{\lambda}{\varepsilon}(1 + \nu x^2)y_2, \quad (35a)$$

$$\frac{dy_1}{dt} = \frac{10}{\varepsilon^2}(y_2 - y_1), \quad (35b)$$

$$\frac{dy_2}{dt} = \frac{1}{\varepsilon^2}(28y_1 - y_2 - y_1y_3), \quad (35c)$$

$$\frac{dy_3}{dt} = \frac{1}{\varepsilon^2}(y_1y_2 - \frac{8}{3}y_3), \quad (35d)$$

where the fast component $y = (y_1, y_2, y_3)^T$ solves the Lorenz equation. In the sequel we investigate two different couplings between the fast and the slow process by choosing $\nu \in \{0, 1\}$.

According to [47, ch. 11.7.2] (see also [19, ex. 6.2]), when eliminating the fast chaotic variable y , the approximate dynamics for $\nu = 0$ is given by

$$dX_t = A(X_t - X_t^3) dt + \sqrt{\sigma} dW_t, \quad (36)$$

with $A = 1$ and the diffusion coefficient that is given by the Green-Kubo formula

$$\sigma = 2\lambda^2 \int_0^\infty \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \psi^s(y) \psi^{s+t}(y) ds dt. \quad (37)$$

Therein $\psi^t(y) = e_2 \cdot \varphi^t(y)$ with $\varphi^t(y)$ denoting the solution of the fast process y at time t when $\varepsilon = 1$ and $e_2 = (0, 1, 0)^T$. The convergence of the solution of (35a) to the solution of (36) can be justified rigorously using the recent results from [42]. However, the above expression for σ is not useful practically: Not only does not it give an analytical value for σ but using it to obtain σ is computationally expensive. Hence, it would be advantageous to use the methodology proposed here and estimate the effective coefficients via observations of the complete (deterministic) fast/slow system. To illustrate numerically that our estimation procedure can indeed deal with this problem, we apply it to observations of the deterministic fast/slow system using $\lambda = 2/45$ and $\nu = 0$ to estimate both the drift and the diffusion coefficients. Since the system is deterministic, classical solvers for ODEs may be employed. For example, depending on the stiffness (i.e. on ε) of the system, either a fourth order Runge-Kutta scheme or a solver based on numerical differentiation formulas is used. Figure 8 depicts the estimated values as functions of the final time $t = nh$ for two different

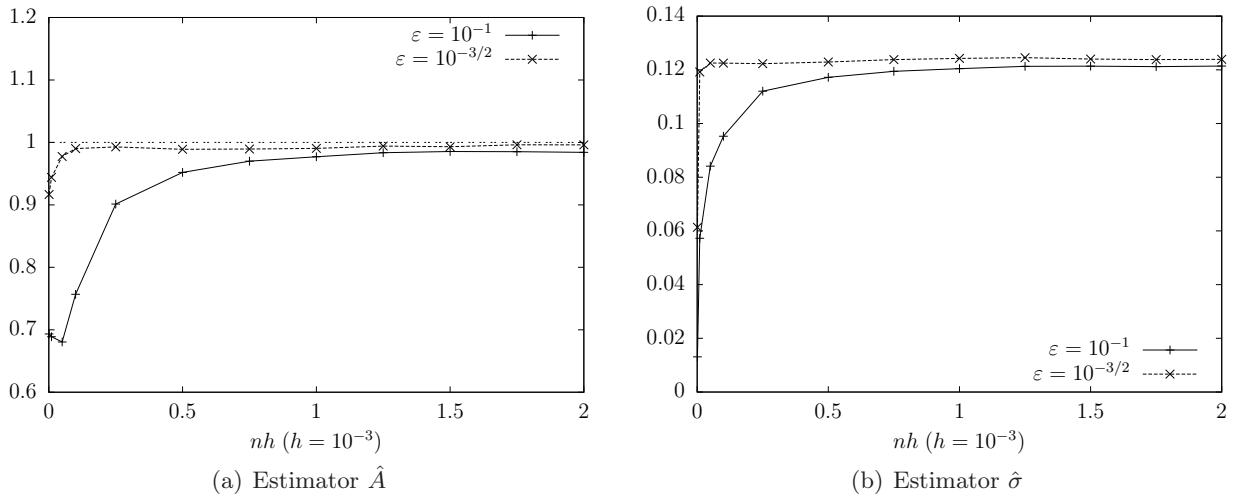


Figure 8: Performance of the estimation scheme applied to the deterministic system (35) using $\lambda = 2/45$, $\nu = 0$, and $\varepsilon \in \{10^{-3/2}, 10^{-1}\}$. The final time of the considered time series is $t = nh$ with $h = 0.001$ and the true effective drift parameter is $A = 1$.

choices of the scale separation $\varepsilon \in \{10^{-3/2}, 10^{-1}\}$. The value $\varepsilon = 10^{-3/2}$ is the same with that used in [19], thus allowing for direct comparisons to be made; we will return to this point shortly. The estimated drift parameter \hat{A} (Figure 8(a)) shows (for both values of ε) the typical behavior we expect in the context of multiscale diffusions. While the estimator is biased for small values of nh , increasing nh significantly reduces the bias so that the estimator approaches the true value (dashed line). The estimators of the effective diffusion coefficient $\hat{\sigma}$ (Figure 8(b)) also approach a limiting value when increasing nh , with minor fluctuations for both values of ε . In both plots, one observes a performance difference of the estimators for different values of ε . In fact, the more distinctive the scale separation (i.e. the smaller ε) between fast and slow components, the faster the estimators approach a limiting value.

For both drift and diffusion estimator the curves for different values of ε give slightly different values

$$\hat{A}_\varepsilon \approx \begin{cases} 0.984 & , \text{ if } \varepsilon = 10^{-1} \\ 0.998 & , \text{ if } \varepsilon = 10^{-3/2} \end{cases} \quad \text{and} \quad \hat{\sigma}_\varepsilon \approx \begin{cases} 0.121 & , \text{ if } \varepsilon = 10^{-1} \\ 0.124 & , \text{ if } \varepsilon = 10^{-3/2} \end{cases}$$

at time $nh = 2$. Hence for $\varepsilon > 0$ there exists an additional bias, which is the reason for the observed difference in the estimators for different ε . On the other hand, one expects that as ε decreases the estimators of the effective coefficients become more accurate, as one observes here for the estimated drift \hat{A} . Thus, $\hat{\sigma}$ is also expected to be more accurate as ε decreases. Since no analytic formula for the diffusion coefficient exists, the estimator $\hat{\sigma}$ is compared with alternative numerical approximations available in [19, ex. 6.2 and ill. 10.5]. In the numerical experiments performed in this study the value $\varepsilon = 10^{-3/2}$ was adopted giving a value of 0.126 ± 0.003 using Gaussian (second) moment approximations based on a modified Euler-Maruyama discretization of the effective dynamics, and a value of 0.13 ± 0.01 based on the heterogeneous multiscale method (HMM) with a discretization of the Green-Kubo formula for the effective diffusion coefficient; see also [14] for more elaborated HMM based numerical schemes applied to the Lorenz 96 model. Thus, we have a very good agreement of the result obtained by the estimation proposed here with these previously reported values. It is worth to mention that, unlike the procedure introduced here, the methods employed to determine the effective diffusion coefficient in [19] assume that the effective drift parameter A is known. While the HMM can easily be adapted to the case of an unknown drift

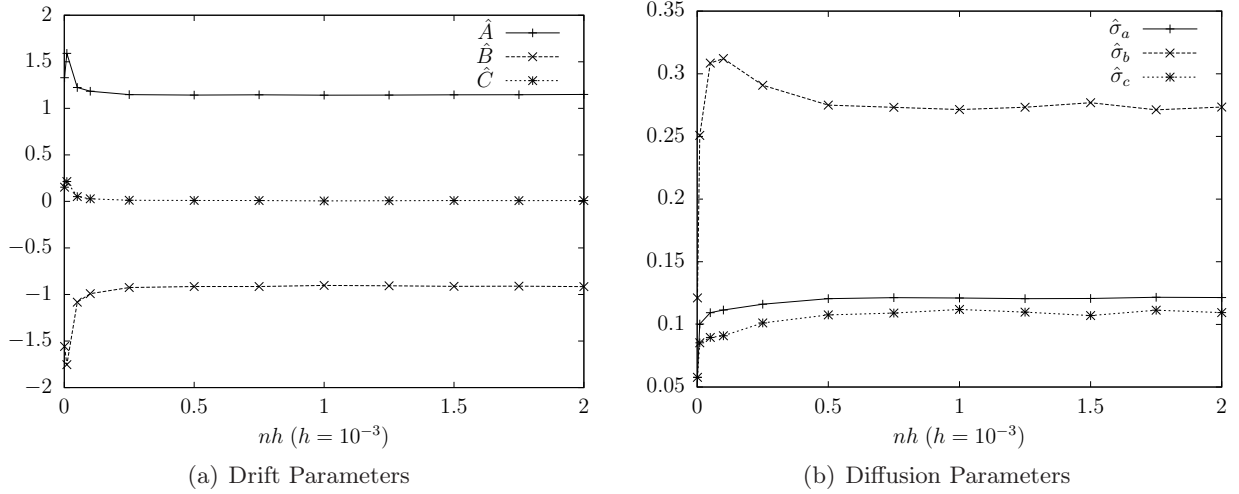


Figure 9: Performance of the estimation scheme applied to the deterministic system (35) using $\lambda = 2/45$, $\nu = 1$, and $\varepsilon = 10^{-3/2}$. The final time of the considered time series is $t = nh$ with $h = 0.001$.

parameter, it is not straightforward to incorporate the unknown drift parameter in the estimation based on Gaussian moment approximations. In any event, incorporating the drift estimation would yield an even larger statistical error for these methods, while the results based on the presented scheme show only minor fluctuations; see Figure 8.

Choosing $\nu = 1$ in (35a) and following the methodology outlined in [47, ch. 11], yields the following effective dynamics

$$dX_t = (AX_t + BX_t^3 + CX_t^5) dt + \sqrt{\sigma_a + \sigma_b X_t^2 + \sigma_c X_t^4} dW_t. \quad (38)$$

The effective coefficients are now given by

$$A = 1 + \sigma, \quad B = \sigma - 1, \quad C = 0,^6 \quad \sigma_a = \sigma, \quad \sigma_b = 2\sigma, \quad \sigma_c = \sigma,$$

with σ being as in (37). We apply our methodology again to observations of the deterministic fast/slow system using $\lambda = 2/45$, $\nu = 1$, and $\varepsilon = 10^{-3/2}$ to estimate all six effective coefficients. Figure 9 illustrates the estimated values of both drift and diffusion parameters as functions of the final time $t = nh$. We observe the procedure's typical behavior in the context of multiscale observations. In fact, for both drift (Figure 9(a)) and diffusion parameters (Figure 9(b)) the estimators approach limiting values with only minor fluctuations when increasing nh .

4 Conclusion

We have developed a numerical methodology for estimating multiple parameters in a coarse-grained equation (in one or multiple dimensions) based on observation from an associated fast/slow system that posses a multiscale structure. This problem is far from straightforward, not only due to the multiscale effects present in the available data, but also due to difficulties associated with estimating parameters when both the drift and the diffusion coefficients are state dependent.

⁶Although we know theoretically that $C = 0$, we estimate C nonetheless with the proposed scheme to illustrate that the novel scheme can correctly identify the relevant parameters in a model that contains more parameters than necessary, for instance given by a Taylor or Fourier series expansion.

The approach developed in this study combines a number of different techniques. On the one hand, the derivation of the estimators relies on simple identities based on the martingale property for stochastic integrals and Itô isometry. On the other, we exploit our freedom in varying the initial condition in combination with standard techniques from inverse problems to define the parameter estimators via best approximation.

We demonstrated via a detailed numerical study that the proposed inference scheme provides us with accurate estimates for parameters in both the drift and the diffusion coefficients in systems with multiscale structure and state dependent noise. In fact, the proposed methodology appears to be accurate and effective even when the stochastic noise in the system is replaced by deterministic chaos.

While the feasibility study of the parameter estimation for multiscale diffusions and the initial presentation of the estimation scheme is the main development here, clearly many open problems and questions remain to be addressed. One such open question is the rigorous analysis of the algorithm to investigate its asymptotic properties and to scrutinize its limitations. Furthermore, the rigorous analysis of the algorithm is expected to reveal insights into the dependency on the algorithm-defining parameters that in turn can be used to reduce the computational complexity of the methodology.

Clearly, from a practical point of view there are different strategies to improve the efficiency of the estimators. A first starting-point could be the usage of techniques with an accelerated convergence instead of the brute-force Monte Carlo sampling to approximate the involved expectations, e.g. quasi Monte Carlo [43] or variance reduction techniques [33, ch. 16]. Also recent work on Multilevel Monte Carlo methods [18] appears very appealing in this context, although care might have to be taken due to the nonlinear nature of the underlying model SDE; cf. [27].

Several questions also arise naturally within the presented framework of varying the initial condition to set up a system of equations: How many initial conditions need to be considered? Where to locate the initial conditions: Equispaced or distributed differently? How does the choice of the initial condition influence the accuracy of the estimator? In fact, preliminary numerical experiments suggest that an alternative distribution of the initial conditions improves the accuracy. Hence, an “optimal” distribution of the initial condition is expected to reduce the computational cost considerably.

Another interesting point concerns the best approximation. Here, we defined the estimator as the element that minimizes the residual of a linear system of equations with respect to the Euclidean norm. Thus, natural questions are, e.g. as to whether an alternative norm might be more appropriate, and how regularizing the minimization problem (e.g. by a truncated singular value decomposition or a Tikhonov regularization) affects the estimator. Also methodologies that “optimize” the linear system of equations to obtain more accurate best approximations, as they are for instance used in the reconstruction of tomographic problems [39], are appealing.

From a computational point of view, employing parallel computing strategies seem also promising. In fact, the presented algorithm consists of multiple parts (mainly when generating the observations) without mutual dependency. Thus, these parts lead to problems that are straightforward to parallelize.

A further important question is whether it is possible to modify the scheme proposed here such that it relies on a single long time-series rather than multiple short simulations using different initial conditions. This would be of interest because in some application areas it is more convenient to provide data of one long simulation rather than multiple short ones. We will examine these and related issues in future studies.

Acknowledgements

We are grateful to the anonymous referees for insightful comments and suggestions. We also thank R. Nürnberg and A. M. Stuart for critically reading an earlier version of the manuscript and for their useful comments and suggestions. This work is supported by the Engineering and Physical Sciences Research Council of the UK through Grant No. EP/H034587.

References

- [1] A. ABDULLE AND G. A. PAVLIOTIS, *Numerical Methods for Stochastic Partial Differential Equations with Multiple Scales*, J. Comput. Phys., 231 (2012), pp. 2482–2497.
- [2] M. ABRAMOWITZ AND I. A. STEGUN, *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables*, Dove, New York, 1964.
- [3] R. AZENCOTT, A. BERI, AND I. TIMOFEYEV, *Adaptive Sub-sampling for Parametric Estimation of Gaussian Diffusions*, J. Stat. Phys., 139 (2010), pp. 1066–1089.
- [4] ———, *Parametric Estimation of Stationary Stochastic Processes under Indirect Observability*, J. Stat. Phys., 144 (2011), pp. 150–170.
- [5] A. BEN-ISRAEL AND T. N. E. GREVILLE, *Generalized Inverses: Theory and Applications*, CMS Books in Mathematics, Springer, 2003.
- [6] D. BLÖMKER, M. HAIRER, AND G. A. PAVLIOTIS, *Multiscale Analysis for Stochastic Partial Differential Equations with Quadratic Nonlinearities*, Nonlinearity, 20 (2007), pp. 1721–1744.
- [7] C. P. CALDERON, *Fitting effective diffusion models to data associated with a "glassy" potential: Estimation, classical inference procedures, and some heuristics*, Multiscale Model. Simul., 6 (2007), pp. 656–687.
- [8] A. CHAUVIÈRE, L. PREZIOSI, AND C. VERDIER, eds., *Cell Mechanics: From Single Scale-Based Models to Multiscale Modeling*, Mathematical & Computational Biology Series, Chapman & Hall/CRC, 2010.
- [9] C. J. COTTER AND G. A. PAVLIOTIS, *Estimating eddy diffusivities from noisy Lagrangian observations*, Commun. Math. Sci., 7 (2009), pp. 805–838.
- [10] M. C. CROSS AND P. C. HOHENBERG, *Pattern formation outside of equilibrium*, Rev. Modern Phys., 65 (1993), pp. 851–1112.
- [11] D. CRUZ-URIBE AND C. J. NEUGEBAUER, *Sharp Error Bounds for the Trapezoidal Rule and Simpson's Rule*, JIPAM. J. Inequal. Pure Appl. Math., 3 (2002), pp. 1–22.
- [12] W. E, D. LIU, AND E. VANDEN-EIJNDEN, *Analysis of multiscale methods for stochastic differential equations*, Comm. Pure Appl. Math., 58 (2005), pp. 1544–1585.
- [13] A. EINSTEIN, *Investigations on the Theory of Brownian Movement*, Dover Publications, 1956.
- [14] I. FATKULLIN AND E. VANDEN-EIJNDEN, *A computational strategy for multiscale systems with applications to lorenz 96 model*, J. Comput. Phys., 200 (2004), pp. 605–638.
- [15] J. FISH, *Multiscale Methods: Bridging the Scales in Science and Engineering*, Oxford University Press, 2009.

- [16] Y. FREDERIX AND D. ROOSE, *Drift-Filtered Approach to Diffusion Estimation for Multiscale Processes*, in *Coping with Complexity: Model Reduction and Data Analysis*, A. N. Gorban and D. Roose, eds., vol. 75 of *Lecture Notes in Computational Science and Engineering*, Springer, 2011, pp. 269–286.
- [17] Y. FREDERIX, G. SAMAEY, AND D. ROOSE, *An analysis of noise propagation in the multiscale simulation of coarse Fokker-Planck equations*, *ESAIM Math. Model. Numer. Anal.*, 45 (2011), pp. 541–561.
- [18] M. B. GILES, *Multilevel Monte Carlo path simulation*, *Oper. Res.*, 56 (2008), pp. 607–617.
- [19] D. GIVON, R. KUPFERMAN, AND A. M. STUART, *Extracting macroscopic dynamics: model problems and algorithms*, *Nonlinearity*, 17 (2004), pp. 55–127.
- [20] G. H. GOLUB AND C. F. VAN LOAN, *Matrix Computations*, *Johns Hopkins Studies in the Mathematical Sciences*, Johns Hopkins University Press, 3rd ed., 1996.
- [21] M. GRIEBEL, S. KNAPEK, AND G. W. ZUMBUSCH, *Numerical Simulation in Molecular Dynamics: Numerics, Algorithms, Parallelization, Applications*, *Texts in Computational Science and Engineering*, Springer, 2007.
- [22] J. D. HAMILTON, *Time Series Analysis*, Princeton University Press, 1994.
- [23] L. P. HANSEN, *Large Sample Properties of Generalized Method of Moments Estimators*, *Econometrica*, 50 (1982), pp. 1029–1054.
- [24] M. F. HORSTEMEYER, *Multiscale Modeling: A Review*, in *Practical Aspects of Computational Chemistry*, J. Leszczynski and M. K. Shukla, eds., Springer, 2010, pp. 87–135.
- [25] W. HORSTHEMKE AND R. LEFEVER, *Noise-Induced Transitions: Theory and Applications in Physics, Chemistry, and Biology*, vol. 15 of *Springer Series in Synergetics*, Springer, 1984.
- [26] P. HUERRE AND M. ROSSI, *Hydrodynamic instabilities in open flows*, in *Hydrodynamic and Nonlinear Instabilities*, C. Godrèche and P. Manneville, eds., Cambridge University Press, 1998, pp. 81–294.
- [27] M. HUTZENTHALER, A. JENTZEN, AND P. E. KLOEDEN, *Divergence of the multilevel Monte Carlo Euler method for nonlinear stochastic differential equations*. <http://arxiv.org/abs/1105.0226>, 2011.
- [28] S. KALLIADASIS, *Nonlinear instability of a contact line driven by gravity*, *J. Fluid Mech.*, 413 (2000), pp. 355–378.
- [29] S. KALLIADASIS, C. RUYER-QUIL, B. SCHEID, AND M. G. VELARDE, *Falling Liquid Films*, vol. 176 of *Applied Mathematical Sciences*, Springer, 2012.
- [30] I. G. KEVREKIDIS, C. W. GEAR, AND G. HUMMER, *Equation-free: The computer-aided analysis of complex multiscale systems*, *AIChE J.*, 50 (2004), pp. 1346–1355.
- [31] I. G. KEVREKIDIS, C. W. GEAR, J. M. HYMAN, P. G. KEVREKIDIS, O. RUNBORG, AND C. THEODOROPOULOS, *Equation-Free, Coarse-Grained Multiscale Computation: Enabling Microscopic Simulators to Perform System-Level Analysis*, *Commun. Math. Sci.*, 1 (2003), pp. 715–762.
- [32] I. G. KEVREKIDIS AND G. SAMAEY, *Equation-Free Multiscale Computation: Algorithms and Applications*, *Annu. Rev. Phys. Chem.*, 60 (2009), pp. 321–344.

- [33] P. E. KLOEDEN AND E. PLATEN, *Numerical Solution of Stochastic Differential Equations*, Springer, 1992.
- [34] N. V. KRYLOV, *On Kolmogorov's equations for finite dimensional diffusions*, in Stochastic PDE's and Kolmogorov equations in infinite dimensions (Cetraro, 1998), vol. 1715 of Lecture Notes in Mathematics, Springer, 1999, pp. 1–63.
- [35] Y. KURAMOTO, *Chemical Oscillations, Waves, and Turbulence*, Chemistry Series, Dover Publications, 2003.
- [36] Y. A. KUTOYANTS, *Statistical Inference for Ergodic Diffusion Processes*, Springer, 2004.
- [37] R. S. LIPTSER AND A. N. SHIRYAEV, *Statistics of Random Processes: I. General Theory*, Stochastic Modelling and Applied Probability Series, Springer, 2nd ed., 2010. Translated by A. B. Aries.
- [38] A. W. LO, *Maximum Likelihood Estimation of Generalized Itô Processes with Discretely Sampled Data*, Econometric Theory, 4 (1988), pp. 231–247.
- [39] H. LUSTFELD, J. A. HIRSCHFELD, M. REISSEL, AND B. STEFFEN, *Enhancement of precision and reduction of measuring points in tomographic reconstructions*, Phys. Lett. A, 375 (2011), pp. 1167–1171.
- [40] A. J. MAJDA, C. FRANZKE, AND B. KHOUIDER, *An applied mathematics perspective on stochastic modelling for climate*, Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci., 366 (2008), pp. 2427–2453.
- [41] R. M. MAZO, *Brownian Motion: Fluctuations, Dynamics, and Applications*, International Series of Monographs on Physics, Oxford University Press, 2002.
- [42] I. MELBOURNE AND A.M.STUART, *A Note on Diffusion Limits of Chaotic Skew Product Flows*, Nonlinearity, 24 (2011), pp. 1361–1367.
- [43] H. NIEDERREITER, *Random Number Generation and Quasi-Monte Carlo Methods*, CBMS-NSF regional conference series in applied mathematics, Society for Industrial and Applied Mathematics, 1992.
- [44] J. NOLEN, G. A. PAVLIOTIS, AND A. M. STUART, *Multiscale Modelling and Inverse Problems*, in Numerical Analysis of Multiscale Problems, O. L. I.G. Graham, T.Y. Hou and R. Scheichl, eds., vol. 83, Springer, 2012, pp. 1–34.
- [45] A. PAPAVALIOU, G. A. PAVLIOTIS, AND A. M. STUART, *Maximum likelihood drift estimation for multiscale diffusions*, Stochastic Process. Appl., 119 (2009), pp. 3173–3210.
- [46] G. A. PAVLIOTIS AND A. M. STUART, *Parameter Estimation for Multiscale Diffusions*, J. Stat. Phys., 127 (2007), pp. 741–781.
- [47] ———, *Multiscale Methods: Averaging and Homogenization*, Springer, 2008.
- [48] R. PENROSE, *On best approximate solutions of linear matrix equations*, Math. Proc. Cambridge Philos. Soc., 52 (1956), pp. 17–19.
- [49] M. PRADAS, G. A. PAVLIOTIS, S. KALLIADASIS, D. T. PAPAGEORGIOU, AND D. TSELUIKO, *Additive noise effects in active nonlinear spatially extended systems*, European J. Appl. Math., (2012). DOI: 10.1017/S0956792512000125.

- [50] M. PRADAS, D. TSELUIKO, S. KALLIADASIS, D. T. PAPAGEORGIOU, AND G. A. PAVLIOTIS, *Noise Induced State Transitions, Intermittency, and Universality in the Noisy Kuramoto-Sivashinsky Equation*, Phys. Rev. Lett., 106 (2011), pp. 060602–1–060602–4.
- [51] B. L. S. PRAKASA RAO, *Statistical Inference for Diffusion Type Processes*, vol. 8 of Kendall's Library of Statistics, Arnold, 1999.
- [52] N. SAVVA, S. KALLIADASIS, AND G. A. PAVLIOTIS, *Two-Dimensional Droplet Spreading over Random Topographical Substrates*, Phys. Rev. Lett., 104 (2010), pp. 084501–1–084501–4.
- [53] C. THEODOROPOULOS, Y.-H. QIAN, AND I. G. KEVREKIDIS, *"Coarse" Stability and Bifurcation Analysis Using Time-Steppers: A Reaction-Diffusion Example*, Proc. Natl. Acad. Sci. USA, 97 (2000), pp. 9840–9843.
- [54] E. VANDEN-EIJNDEN, *Numerical techniques for multi-scale dynamical systems with stochastic effects*, Commun. Math. Sci., 1 (2003), pp. 385–391.
- [55] L. ZHANG, P. A. MYKLAND, AND Y. AÏT-SAHALIA, *A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data*, J. Amer. Statist. Assoc., 100 (2005), pp. 1394–1411.
- [56] R. ZWANZIG, *Nonequilibrium Statistical Mechanics*, Oxford University Press, 2001.